ProteinLynx Global SERVER Version 2.2.5

User's Guide

71500125602 / Revision A

Copyright © Waters Corporation 2006.
All rights reserved.

Copyright notice

© 2006 WATERS CORPORATION. PRINTED IN THE UNITED STATES OF AMERICA AND IRELAND. ALL RIGHTS RESERVED. THIS DOCUMENT OR PARTS THEREOF MAY NOT BE REPRODUCED IN ANY FORM WITHOUT THE WRITTEN PERMISSION OF THE PUBLISHER.

The information in this document is subject to change without notice and should not be construed as a commitment by Waters Corporation. Waters Corporation assumes no responsibility for any errors that may appear in this document. This document is believed to be complete and accurate at the time of publication. In no event shall Waters Corporation be liable for incidental or consequential damages in connection with, or arising from, its use.

Waters Corporation
34 Maple Street
Milford, MA 01757
USA

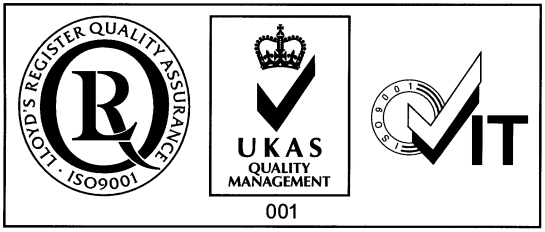
Trademarks

Millennium and Waters are registered trademarks of Waters Corporation. MassLynx and ProteinLynx Global SERVER are trademarks of Waters Corporation.

Windows is a registered trademark of Microsoft Corporation. IBM and AIX are registered trademarks of International Business Machines Corporation. UNIX is a registered trademark of The Open Group. Sun and Solaris are registered trademarks of Sun Microsystems, Inc. Linux is a registered trademark of Linus Torvalds. SUSE is a registered trademark of Novell, Inc. Red Hat is a registered trademark of Red Hat, Inc.

ICAT is a trademark of the University of Washington. iTRAQ is a trademark of Applera Corporation.

Other trademarks or registered trademarks are the sole property of their respective owners.



Intended use

ProteinLynx Global SERVER can be used as a research tool to deliver qualitative protein identification and relative quantification. It is not for use in diagnostic procedures.

Customer comments

Please contact us if you have questions, suggestions for improvements, or find errors in this document. Your comments will help us improve the quality, accuracy, and organization of our documentation.

You can reach us at tech_comm@waters.com.

Table of Contents

1 Installing ProteinLynx Global SERVER 1-1

Typical client/server installation 1-2

Installing PLGS on Windows® 1-3

Backing up the PLGS folders 1-3

Backing up databanks 1-3

Uninstalling PLGS in Windows 1-3

Installing PLGS on Windows 1-4

Restoring backed-up folders 1-5

Running PLGS on Windows in a client/server environment 1-5

Running PLGS on Windows on a single PC 1-6

Starting modules manually and troubleshooting problems 1-6

Installing PLGS on Linux 1-7

Before installing PLGS 1-7

Backing up the PLGS folders 1-7

Backing up databanks 1-7

Changing file permissions 1-8

Uninstalling previous versions of PLGS in Linux 1-8

Installing PLGS on Linux 1-9

Restoring backed-up folders 1-11

Running PLGS on Linux 1-11

Starting modules manually and troubleshooting problems 1-13

Installing PLGS on UNIX 1-15

Before installing PLGS on UNIX 1-15

Backing up the PLGS directory 1-15

Uninstalling a previous version of PLGS 1-15

Installing PLGS on UNIX 1-16

Configuring PLGS on UNIX 1-17

Search engine memory allocation 1-17

TMPDIR environment variable 1-18

Search engine temporary directory 1-18
Running PLGS on UNIX 1-19
Starting modules manually and troubleshooting problems 1-19
Installation troubleshooting on UNIX 1-20
Installer startup problems 1-20
Microkernel failures 1-20
Search engine failures 1-21
Large databank (>2 GB) problems 1-21
Databank and BLAST searching problems 1-21

Restoring old databanks 1-23

Setting the number of processors 1-24
DDA data processing 1-24
Expression data processing 1-25
Databank searching 1-25

2 Setting up ProteinLynx Global SERVER 2-1

ProteinLynx browser 2-2
Tool tray 2-3
Adding and removing tools 2-4

Changing preferences 2-5
Search Engine tab 2-5
Adding a search engine 2-6
Modifying a search engine 2-7
Removing a search engine 2-8
Processors tab 2-8
Adding a processor 2-8
Modifying a processor 2-9
Removing a processor 2-9
Instrument tab 2-10
Bookmarks tab 2-11
Adding a bookmark 2-11
Modifying a bookmark 2-12
Removing a bookmark 2-12

Colours tab 2-12
Setting confidence levels and colors 2-14
Printing tab 2-16

Setting Automation Setup parameters 2-18
Parameters tab 2-18
Spectrum Output tab 2-20
PlugIns tab 2-23
Replacing the Import PlugIn or adding an Export PlugIn 2-24
Modifying an Export PlugIn 2-27
Removing an Export PlugIn 2-28

3 Creating, importing, and managing projects 3-1

Creating a new project 3-2

Importing and exporting projects 3-3

Opening and updating projects 3-5
Updating projects 3-5

Closing and deleting projects 3-6

4 Annotating and tracking samples with Sample Manager 4-1

Getting started with Sample Manager 4-2
Adding a sample 4-2
Deleting a sample 4-2

Sample editor 4-3

Generating processed samples 4-5

5 Specifying samples, vials, and plates with Container Manager .. 5-1

What is Container Manager? 5-2
Workflow templates and Processing parameters 5-2

Importing and viewing PLGS sample lists 5-3
Importing PLGS sample lists 5-3
Sample list requirements 5-4
Viewing PLGS sample lists 5-5

View column 5-7
Processing and Searching 5-7
Changing Templates 5-7

Creating a new vial, microtitre or target plate 5-9

Setting a sample 5-11

Attaching raw data 5-13
Selecting more than one well or spot 5-14

Processing raw data 5-17
Workflow and spectrum icons in the navigator tree 5-18
Viewing the mass spectrum 5-19

Re-searching processed data 5-20

Adding processing parameters templates 5-21

Exporting and importing mass spectra 5-22
Exporting mass spectra 5-22
Importing mass spectra 5-22

Working with plates 5-23
Merging MSMS spectra and results 5-24
Customizing the plate view 5-25

Simplifying peaks with SuperTrack 5-26
Exporting SuperTrack results as XML 5-28

Interfacing with MassLynx 5-29
Exporting a sample list to MassLynx 5-29
Acquiring data 5-31

Troubleshooting failed client-server workflows 5-33

6 Viewing results in the Results Browser 6-1

Viewing results 6-2

Results browser 6-3
Results tree toolbar 6-4
Bottom toolbar 6-5

Spectrum viewer toolbar 6-6
Results browser navigator tree 6-7
Protein view 6-7
Peptide view 6-9
Selecting items in the navigator tree 6-9
PepGrab 6-11
Protein and EST table 6-12
Peptide table 6-13
Controlling the columns in the tables 6-14
Selecting proteins and ESTs from the table 6-15
Selecting peptides from the table 6-15
Resubmitting the search 6-15
Copying data 6-16
Printing the results 6-16
Spectrum Viewer for MS data 6-16
Viewing raw data 6-18
Changing the x-axis view 6-20
Viewing the fragment ion display 6-20
Spectrum Viewer for MSMS data 6-21
Displaying ion probabilities 6-22
Spectrum Viewer options 6-24
Copying data 6-26

Protein Workpad 6-27
Coverage map 6-28
Running a simulated digest 6-29
Retrieving databank entries 6-30

Exclude Masses Workpad 6-31
Adding items to the excluded list 6-32
Deleting items from the excluded list 6-33
Running a simulated digest for a protein 6-33
Viewing the masses associated with an excluded item 6-34

7 Defining templates for searching with Workflow Designer 7-1

What is Workflow Designer? 7-2
The Workflow Designer interface 7-2
Workflow Designer toolbar 7-4

Creating a workflow template 7-5
Editing workflow templates 7-9
Opening workflow templates 7-10

Filters 7-11
AutoMod filter 7-11
De Novo filter 7-11

8 Creating custom processing parameters 8-1

Getting started with the Data Preparation tool 8-2

Attribute sets for data preparation 8-5
MALDI PSD MX 8-5
MALDI Q-Tof MSMS 8-5
Electrospray DDA (QTOF-MSMS) 8-6
Mass Accuracy attributes 8-6
Noise Reduction attributes 8-9
Deisotoping and Centroiding attributes 8-12
Peak Matching attributes 8-15
Chromatogram attributes 8-15

9 Viewing and processing gel data with Gel Manager 9-1

Getting started with Gel Manager 9-2

Adding and importing data 9-3
Adding a new gel without an image 9-3
Importing gel spots 9-3
Importing a gel from an OLB file 9-5
Importing a gel from sample list 9-6
Replacing the sample in a well or spot 9-7

Processing data 9-8

Viewing gel data 9-9
Viewing a gel image 9-9
Viewing a summary of results for a gel 9-9
Viewing sample annotation 9-10

10 Using Expression Analysis to compare and analyze sample groups 10-1

Getting started with Expression Analysis 10-2
Opening a project 10-2

Experiment Analysis Design Manager 10-3
Experiment Attributes 10-4
Select Grouping Method 10-5
Manually Define Experiment Variables 10-6
Manually Assign Samples To Groups 10-7
Select Data 10-7
Assess Data Quality 10-8
Quantitation Analysis 10-8
Starting an Expression analysis 10-9

Viewing Expression Results 10-10
EMRT table 10-10
Protein table 10-13
Filtering the results 10-13
Replicate filter 10-14
Confidence Limit, P value, and Ratio filters 10-15
Additional Filter settings 10-15
Importing workflows 10-16
Searching EMRTs from the EMRT table 10-17

Log Plot Viewer 10-18

Expression Data Viewer 10-20
Group level 10-21
Sample level 10-21
Replicate/Spectrum level 10-21

Exporting Switch Lists **10-23**

Importing Significant Clusters **10-24**
Significant clusters list file format 10-24

Assess Data Quality viewer **10-25**

11 Creating print templates and printing project data **11-1**

Printing data **11-2**

Using print wizards **11-3**
Project print wizard 11-3
Workflow print wizard 11-6

Opening and deleting print templates **11-12**

Creating print templates **11-13**
Adding content to the results nodes 11-15
Filtering, sorting and limiting in results nodes 11-16
Filtering results 11-16
Sorting results 11-17
Limiting results 11-17

Customizing print templates **11-19**
Buttons for adding content to pages 11-23

12 Managing modifier and digest reagents **12-1**

Getting Started with the Modifier tool **12-2**

Viewing existing modifier reagents **12-3**

Adding and editing custom modifier reagents **12-4**
Deleting custom modifier reagents 12-6

Getting started with the Digest Reagent tool **12-7**

Viewing existing digest reagents **12-8**

Custom digest reagents **12-9**
Adding or editing custom digest reagents 12-9
Saving custom digest reagents 12-10

Deleting custom digest reagents 12-10

13 Organizing databanks with the Databank Admin tool 13-1

Getting started with the Databank Admin tool 13-2

Adding databanks 13-3
Databank attributes 13-4

Editing databanks 13-11

Removing and deleting databanks 13-13
Removing databanks from the system record 13-13
Deleting databanks 13-13
Deleting archive files 13-14
Deleting revived archives 13-14
Keeping archived copies of a databank 13-15
Reviving an archive 13-15

Connecting to a search engine 13-17

14 Query Tools 14-1

Query toolbar 14-2

Databank Search tool 14-3
Databank search parameters 14-5
Search Engine Type 14-5
Mass Spectrum (PLGS) or Data File (MASCOT) 14-5
Databanks (PLGS) or Database (MASCOT) 14-6
Species (PLGS) or Taxonomy (MASCOT) 14-6
Peptide Tolerance 14-6
Fragment Tolerance (PLGS) or MSMS Tolerance (MASCOT) 14-7
Estimated Calibration Error (Da or ppm) 14-7
Molecular Weight Range (PLGS) or Protein Mass (MASCOT) 14-8
pI Range 14-8
Minimum Peptides to Match 14-9
Maximum Hits to Return 14-9
Primary Digest Reagent (PLGS) or Enzyme (MASCOT) 14-9

Secondary Digest Reagent 14-10
Missed Cleavages 14-10
Fixed Modifications 14-10
Variable Modifications 14-11
Exclude Masses 14-11
Validate Results 14-12
Monoisotopic or Average 14-12
Mass Values 14-12
Peptide Charge 14-12
Instrument Type 14-13

AutoMod Analysis tool 14-14
AutoMod Analysis search parameters 14-16
Consider Modifications 14-16
Consider Substitutions 14-16
Specifying the maximum substitutions and modifications per peptide 14-16
Specifying the likelihood of substitutions 14-17
Validate Results 14-17
Selecting protein sequences for the search 14-18
Selecting EST sequences for the search 14-18

De Novo Sequencing tool 14-19
De Novo sequencing parameters 14-21
Specifying the estimated calibration error 14-21
Specifying maximum hits to return 14-21
Specifying modifications to peptides 14-21
Validate Results 14-22

BLAST Searching tool 14-23
BLAST search parameters 14-24
Peptide sequence 14-25
Scoring matrix 14-25
Expect Threshold 14-25
Gapped 14-26
Low Complexity Filter 14-26
Number of Hits 14-26

BLAST results 14-26
Navigating within a BLAST results panel 14-27

15 Real Time Databank Searching 15-1

Using real time databank searching 15-2
Launching the Real Time Databank Searching application 15-2
Processing parameters 15-4
Searching parameters 15-5
Real time status 15-7
Setting up a real time databank searching acquisition 15-8
Setting up your DDA file 15-10
De-isotope peak detection 15-11
Tolerance window 15-12
Extraction window 15-12
Exclude window 15-13
Other DDA experiment settings 15-13

Advanced options 15-14
Data processing 15-14
Remote searching 15-14
Displaying diagnostics 15-15

16 Using MS^E for qualitative proteomics 16-1

What is MS^E? 16-2

Creating an MS^E method file 16-3

Running an MS^E experiment 16-7
Necessary sample list fields 16-7

A Quick Start Tutorials A-1

Creating a project and processing acquired data files A-2
Setting samples A-2
Setting the target plate A-2

MALDI test procedure A-5
Setting the target plate A-5

Setting processing parameters A-6
Creating a workflow A-7
Attaching the data processing parameters A-8
Attaching the workflow file A-9
Exporting the sample list to MassLynx A-9
Acquiring data A-11

Acquiring Q-Tof MSMS data A-14
Setting the microtitre plate A-14
Setting processing parameters A-14
Creating a workflow A-17
Attaching the data processing parameters A-18
Attaching the workflow file A-19
Exporting the sample list to MassLynx A-19
Acquiring data A-21

Adding a new databank A-25

B Scoring Schemes B-1

Scoring summary B-2

MALDI scoring (PMF, PMF + fragment ion searches) B-4

MSMS scoring (fragment ion searches) B-5

How do I know if a hit is real? B-6

Automatic data curation B-7
PMF B-7
PMF + Fragment Ion B-7
Fragment Ion B-8
Electrospray-MS B-8
Electrospray-High/Low B-8

C Implementing a plugin for ProteinLynx Global SERVER C-1

An introduction to the PLGS plugin C-2

Plugin architecture C-3

Use case – the PLGS FileSystemPlugIn C-5

XML communication with the plugin implementation C-6

Adding a plugin to the PLGS application C-7

An example Executable plugin C-11

An example Java plugin C-13

Basic plugin-Specific Queries C-16

Selection of elements C-16

Selecting a Project document for a given Project ID C-16

Update of elements C-17

Updating a Project document for a given Project_ID C-18

Deletion of elements C-18

Deleting a Mass Spectrum document for a given Sample Tracking ID . C-18

Insertion of documents C-19

Inserting a Workflow document and updating the associated Project document C-19

Query tag definitions in the ProteinLynx DTD C-21

Plugin process exit codes C-26

UML Class Diagram for the PLGS plugin Architecture C-27

D UNIX Help for Installing PLGS on AIX Platforms D-1

Installing PLGS using the command line D-2

Adding TMPDIR D-4

Mounting a CD-ROM D-4

Using SMIT D-6

Using navigation and installation commands D-8

Creating and managing user accounts and groups D-9

E Databanks – Formats E-1

URL addresses E-2

SPTREMBL flat file format E-3

Genbank flat file format E-6

BLAST flat file format E-8

FASTA flat file format E-9

FASTA STANDARD E-9
FASTA NCBI_EXPASY_STANDARD E-9
FASTA NCBI_PRF_PIR E-10
FASTA NCBI_PDB E-10
FASTA NCBI_PATENT E-11
FASTA NCBI_GENINFO E-11
FASTA NCBI_GENERAL E-11
FASTA NCBI_LOCAL E-11
FASTA PDB E-12
FASTA PIR E-12
FASTA SRS E-13
FASTA ARABIDOPSIS_GENOME E-13
FASTA NRDB E-14
FASTA UNIGENE E-14
FASTA STANDARD_SPACED E-14
FASTA LONG_DESCRIPTION E-15
FASTA ACCESSION_ONLY E-15

Index Index-1

1 Installing ProteinLynx Global SERVER

ProteinLynx™ Global SERVER (PLGS) is a multi-platform Java™, C, and C++ application, which features a new and comprehensive range of integrated tools for proteomics project management, protein quantification, and protein identification and characterization, through exploiting the specificity of exact mass data.

ProteinLynx Global SERVER can be run in a client/server environment, or on a single PC. When run on Linux® or UNIX®, ProteinLynx browser contains the Database Admin Tool and Help.

This chapter describes the procedure for installing PLGS on the following platforms. Each package has its own start-up procedure.

See also: Additional platform-specific information on installation and configuration issues can be found in the ProteinLynx Global SERVER 2.2.5 Release Notes.

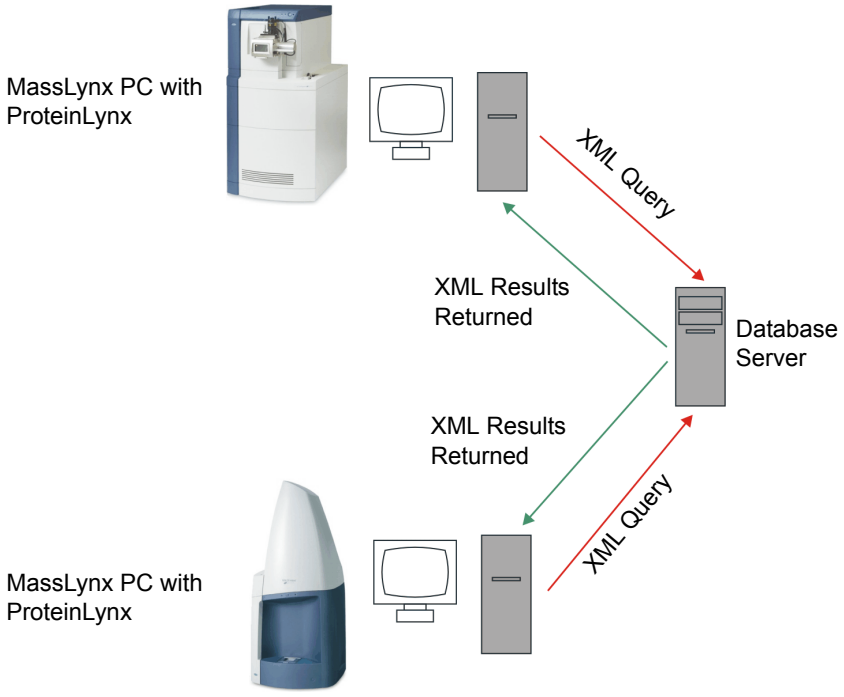
Contents:

Topic	Page
Typical client/server installation	1-2
Installing PLGS on Windows®	1-3
Installing PLGS on Linux	1-7
Installing PLGS on UNIX	1-15
Restoring old databanks	1-23
Setting the number of processors	1-24

Typical client/server installation

The following graphic shows how ProteinLynx Global SERVER is typically used in a client/server environment.

ProteinLynx Global SERVER in a client/server environment:



Installing PLGS on Windows®

This section describes the steps to install and run PLGS on Windows on a single PC or in a client/server environment. However, if you have a previous version of PLGS already installed on your PC, you must:

- back up the PLGS folders.
- back up any databanks that are stored in the installation directory.
- uninstall previous versions of PLGS.

Backing up the PLGS folders

Before uninstalling a previous version PLGS, make a backup copy of the following folders from your PLGS installation directory:

- docs – contains workflow template files, processing parameters files, and so on.
- root – contains project files that you have created.

Backing up databanks

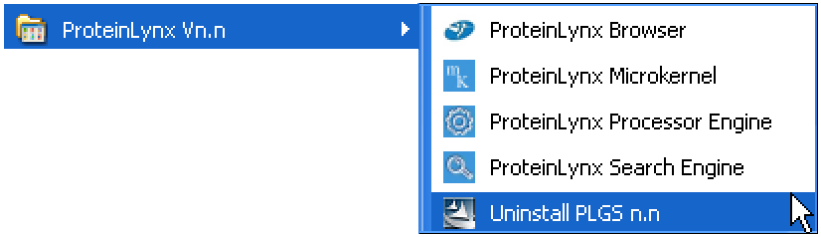
If any of your databanks are stored in the directory in which PLGS is installed, you must make backups of the databanks before uninstalling PLGS.

Uninstalling PLGS in Windows

To uninstall a previous version of PLGS:

1. From the ProteinLynx program group, select the uninstall option.

ProteinLynx program group - uninstall option:



Exception: If you are uninstalling PLGS 2.2.5, the Microkernel, Processor Engine, and Search Engine options are not displayed in the program group.

2. Follow the instructions in the Uninstaller wizard.

Installing PLGS on Windows

To install PLGS on Windows:

1. Double-click the PLGS2.2.5_WINDOWS.exe file to open the InstallShield Wizard.

   Result: After a short pause, the ProteinLynx Global SERVER installation wizard will be displayed.

2. Click Next.

3. Read and understand the terms of the license agreement, select the accept option, and then click Next.

4. In the product destination screen, do one of the these actions:

   - Click Next to accept the default installation location (C:\PLGS<version number>).
   - Browse for another directory, and then click Next.

5. If the installer cannot detect a valid IP address or if it detects multiple IP addresses, the Specify IP Address screen is displayed.

   Rule: If the installer detects a valid IP address, this screen is not displayed.

   Type the IP address of the network connection, and then click Next. If you cannot identify the IP address, ask your system administrator for help in doing so.

6. On the Install as Services screen, select whether you want to install as services:

   - Yes – The search engine and processor automatically run in the background when the PC is running. Data on mapped drives cannot be processed or searched if the modules are run as services.
   - No (default) – The search engine and processor only start when you start the ProteinLynx Browser.

   Recommendation: Select No if you are running PLGS on a single PC.

7. Click Next, and then review the installation summary information. If you wish to change any of the options, click Back. If you are ready to install, click Install.

   Tip: Once the installation starts, it can be stopped by clicking Cancel.

Once the installation is complete the Installation Successful dialog box is displayed. Click Finish to close the Installer. The ProteinLynx program group is now available.

ProteinLynx program group:

Restoring backed-up folders

If you uninstalled a previous version of PLGS and backed-up folders (see Backing up the PLGS folders on page 1-3), you should restore them before starting PLGS. To do this, copy the backed-up docs and root folders into the folder where you installed PLGS.

If you backed up databanks, they must be re-added to PLGS. For details on how to do this, see Adding databanks on page 13-3.

Running PLGS on Windows in a client/server environment

To run PLGS in a client/server environment you need to start these PLGS modules on each computer:

- Microkernel
- Search engine
- Processor

All of these modules are started automatically when you start the ProteinLynx browser on that computer.

To start the PLGS browser:

1. Click Start > All Programs > ProteinLynx > ProteinLynx Browser.

Running PLGS on Windows on a single PC

To start ProteinLynx Global SERVER, click Start > All Programs > ProteinLynx > ProteinLynx Browser.

Starting modules manually and troubleshooting problems

All of the modules on a computer are started automatically when you start the ProteinLynx browser. Nevertheless, you might wish to start the individual modules separately.

To start PLGS modules manually:

1. Navigate to the PLGS installation directory, and then to the bin subdirectory.
2. Start the module by double-clicking in Windows, or by typing its name at the command prompt.
   - ProcessorEngine.bat to start the processor.
   - SearchEngine.bat to start the search engine.
   - PLmicrokernel.exe to start the microkernel.

If you start the modules automatically, by starting the ProteinLynx browser, log files are generated by the software. These log files can help you to solve operational problems, and will be helpful to Waters if you request technical support.

To view log files:

1. Navigate to the PLGS installation directory, and then to the log subdirectory.
2. Open the log file in a text editor, such as Notepad. Two log files are created:
   - Processor.txt for the processor log.
   - SearchEngine.txt for the search engine and microkernel log.

Installing PLGS on Linux

This section describes the steps required to install and run PLGS on Linux. PLGS can be installed under Red Hat® Linux 9 on Intel-based architectures, or SUSE® Linux Enterprise Server 9 on IBM Power architectures.

On Linux, the ProteinLynx browser enables you to add new databanks to the server, or view online help (see Linux ProteinLynx browser: on page 1-13).

Restriction: Only the Databank Admin Tool and the online Help are available in the Linux PLGS browser. If so configured, processing and searching can be run on a Linux machine from a remote Windows PLGS browser.

Rule: All UNIX commands are case sensitive.

Before installing PLGS

Complete these tasks before installing PLGS in Linux:

- Back up the PLGS directories (see Backing up the PLGS folders on page 1-7).
- Ensure that you are logged on with root permissions (see Changing file permissions on page 1-8).
- Uninstall previous versions of PLGS (see Uninstalling previous versions of PLGS in Linux on page 1-8).

Backing up the PLGS folders

Before installing PLGS, make a backup copy of the following folders:

- docs
- root

Backing up databanks

If any of your databanks are stored in the directory in which PLGS is installed, you must make backups of the databanks before uninstalling PLGS.

Changing file permissions

File permissions exist on Linux to prevent unauthorized access. Before installing PLGS, ensure you are logged on with user ROOT permissions. If file permissions problems continue, you need to change the file permissions.

To change a file's permissions:

1. Log on as the root user.
2. Use the cd command to navigate to the file's folder.
3. Change the file's permission settings by typing:

   `chmod 777 [filename]`

   This removes the restrictions on all file permissions.

Uninstalling previous versions of PLGS in Linux

Previous version of PLGS can be uninstalled from a command prompt or by using the GUI. The uninstaller deletes all folders and contents that were installed with PLGS, and any folders and files that you created using PLGS.

To uninstall PLGS using the command prompt:

1. Open a terminal window and type:

   `cd [PLGS_INSTALL_FOLDER]/_uninst/`

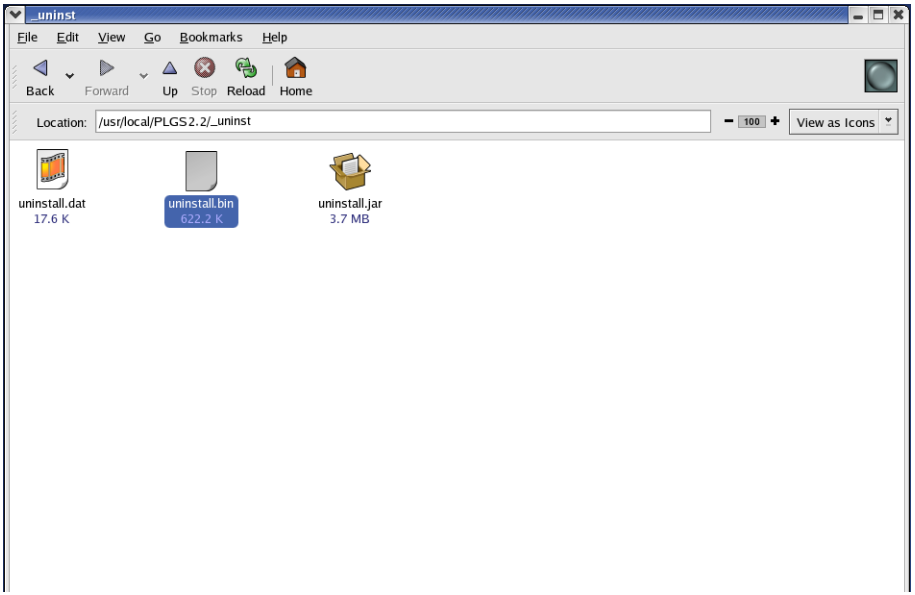
   This takes you to the uninstall folder.
2. To run the uninstaller program, type:

   `./uninstall.bin`
3. Follow the instructions in the Uninstaller wizard.

To uninstall PLGS using the graphical user interface (GUI):

1. Navigate to the `[_uninstall]` folder.

_uninstall folder:



2. Double-click uninstall.bin
3. Follow the instructions in the Uninstaller Wizard.

Installing PLGS on Linux

PLGS can be installed from a command prompt or by using the graphical user interface (GUI). Linux will automatically detect when you load the installation CD.

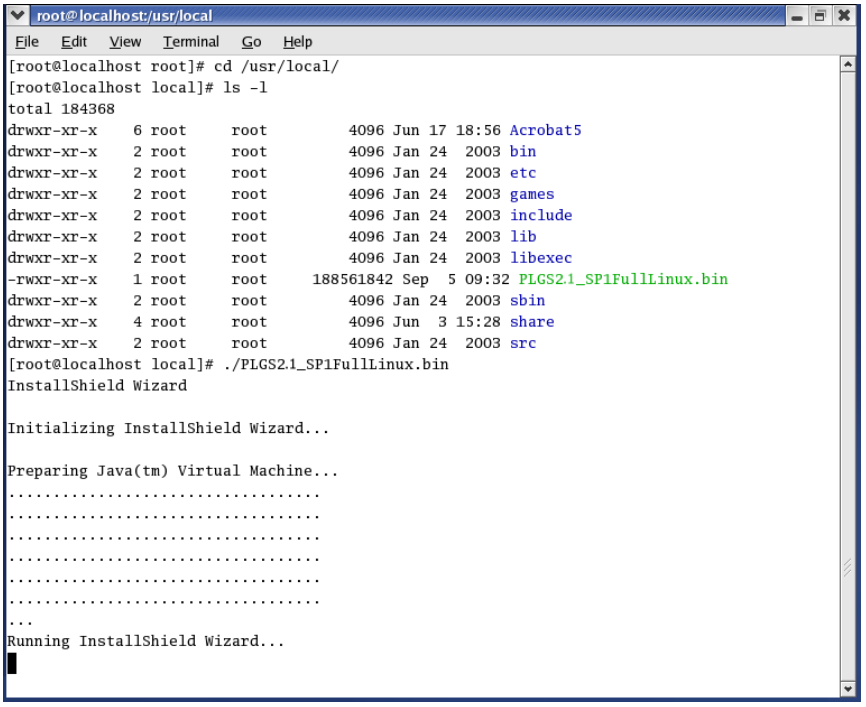
Requirements: If you are installing on SUSE Linux, you must ensure that the IBM C++ Runtime Libraries are installed and that the Java JIT compiler is turned off. For further assistance, refer to the ProteinLynx Global SERVER Release Notes.

To install PLGS from a command prompt:

1. Open a terminal window and navigate to the installation directory using the command

   ```
   cd /usr/local/
   ```

Running InstallShield:


root@localhost:/usr/local
File Edit View Terminal Go Help
[root@localhost root]# cd /usr/local/
[root@localhost local]# ls -l
total 184368
drwxr-xr-x 6 root root 4096 Jun 17 18:56 Acrobat5
drwxr-xr-x 2 root root 4096 Jan 24 2003 bin
drwxr-xr-x 2 root root 4096 Jan 24 2003 etc
drwxr-xr-x 2 root root 4096 Jan 24 2003 games
drwxr-xr-x 2 root root 4096 Jan 24 2003 include
drwxr-xr-x 2 root root 4096 Jan 24 2003 lib
drwxr-xr-x 2 root root 4096 Jan 24 2003 libexec
-rwxr-xr-x 1 root root 188561842 Sep 5 09:32 PLGS2.1_SP1FullLinux.bin
drwxr-xr-x 2 root root 4096 Jan 24 2003 sbin
drwxr-xr-x 4 root root 4096 Jun 3 15:28 share
drwxr-xr-x 2 root root 4096 Jan 24 2003 src
[root@localhost local]# ./PLGS2.1_SP1FullLinux.bin
InstallShield Wizard
Initializing InstallShield Wizard...
Preparing Java(tm) Virtual Machine...
Running InstallShield Wizard...

Tip: Use the ls –l command to list all the files and directories – and their current permissions – in the current directory.

2. Run the binary file using the command:

   `./PLGS2.2.5_INTEL_LINUX.bin`

   or, for SUSE Linux systems:

   `./PLGS2.2.5_PPC_LINUX.sh`

   Result: The ProteinLynx Installer dialog box opens.

3. Specify or browse for a directory in which to install PLGS.

   Recommendation: Install the PLGS in the directory /usr/local/. The default directory is /usr/local/PLGS2.2.5.

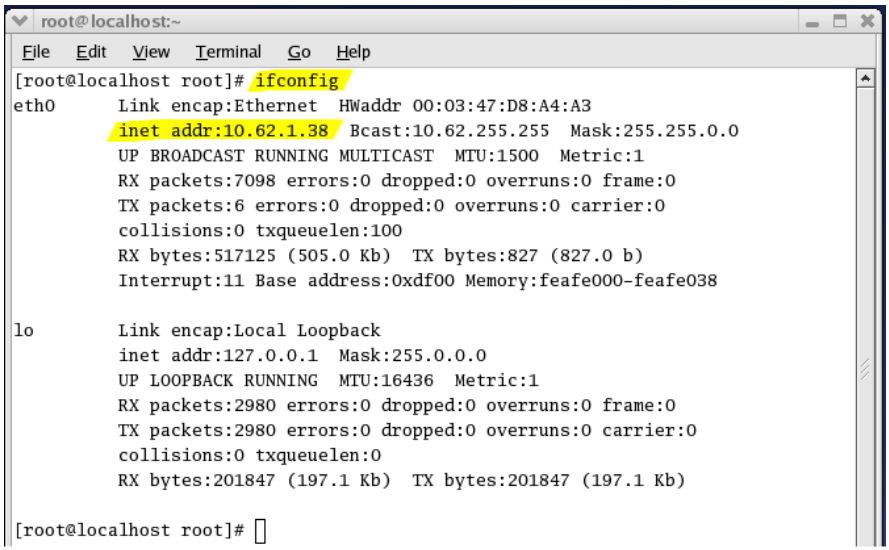
4. Specify the computer's IP address. If needed, use the `ifconfig` command to find the IP address:

   1. Open a terminal window.

2. In the terminal window, type:

   `ifconfig`

ifconfig command:


root@localhost:~
File Edit View Terminal Go Help
[root@localhost root]# ifconfig
eth0 Link encap:Ethernet HWaddr 00:03:47:D8:A4:A3
inet addr:10.62.1.38 Bcast:10.62.255.255 Mask:255.255.0.0
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
RX packets:7098 errors:0 dropped:0 overruns:0 frame:0
TX packets:6 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:100
RX bytes:517125 (505.0 Kb) TX bytes:827 (827.0 b)
Interrupt:11 Base address:0xdf00 Memory:feafe000-feafe038
lo Link encap:Local Loopback
inet addr:127.0.0.1 Mask:255.0.0.0
UP LOOPBACK RUNNING MTU:16436 Metric:1
RX packets:2980 errors:0 dropped:0 overruns:0 frame:0
TX packets:2980 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:0
RX bytes:201847 (197.1 Kb) TX bytes:201847 (197.1 Kb)
[root@localhost root]#

The IP address is displayed on the line `inet addr`.

5. Click Next.The PLGS Installer program starts.

Restoring backed-up folders

If you uninstalled a previous version of PLGS and backed-up folders (see Backing up the PLGS folders on page 1-7), you should restore them before starting PLGS. To do this, copy the backed-up docs and root folders into the folder where you installed PLGS.

If you backed up databanks, they must be re-added to PLGS. For details on how to do this, see Adding databanks on page 13-3.

Running PLGS on Linux

To run PLGS you need to start these PLGS modules on each computer:

- Search engine
- Microkernel

- Processor

These modules are started automatically when you start the PLGS browser on the machine.

PLGS can be run from a command prompt or by using the GUI.

To run PLGS using the command prompt:

1. Open a terminal window, and then type

   `cd <PLGS install location>/bin`

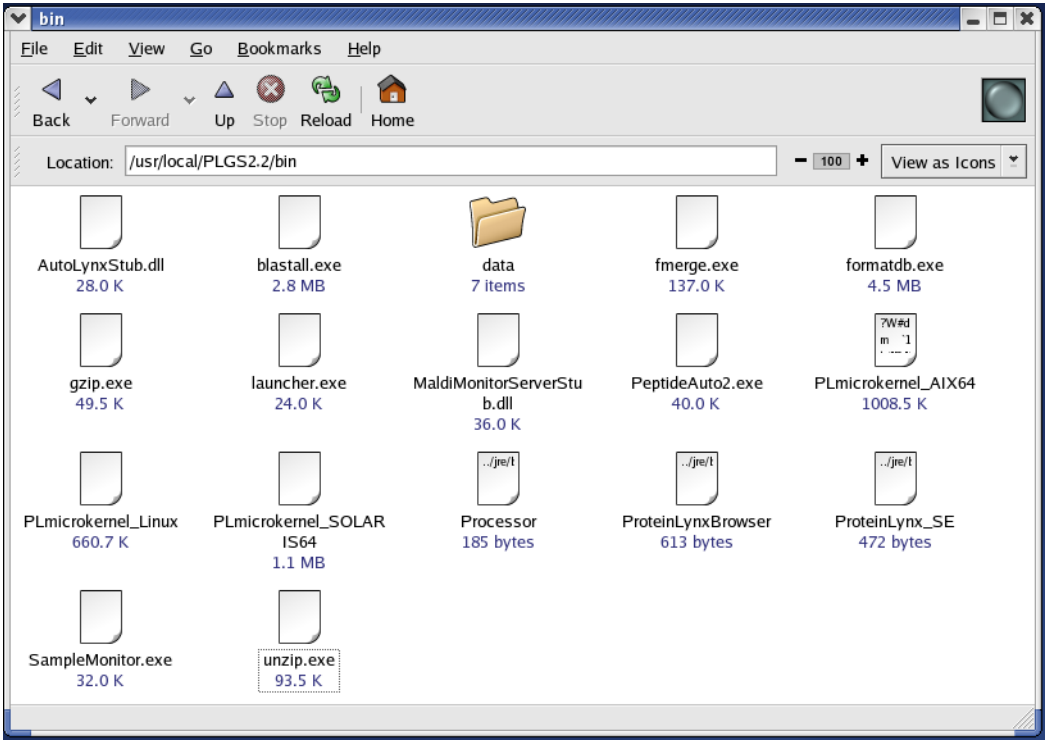
2. To start the browser, type

   `./ProteinLynxBrowser`

To start PLGS using the GUI:

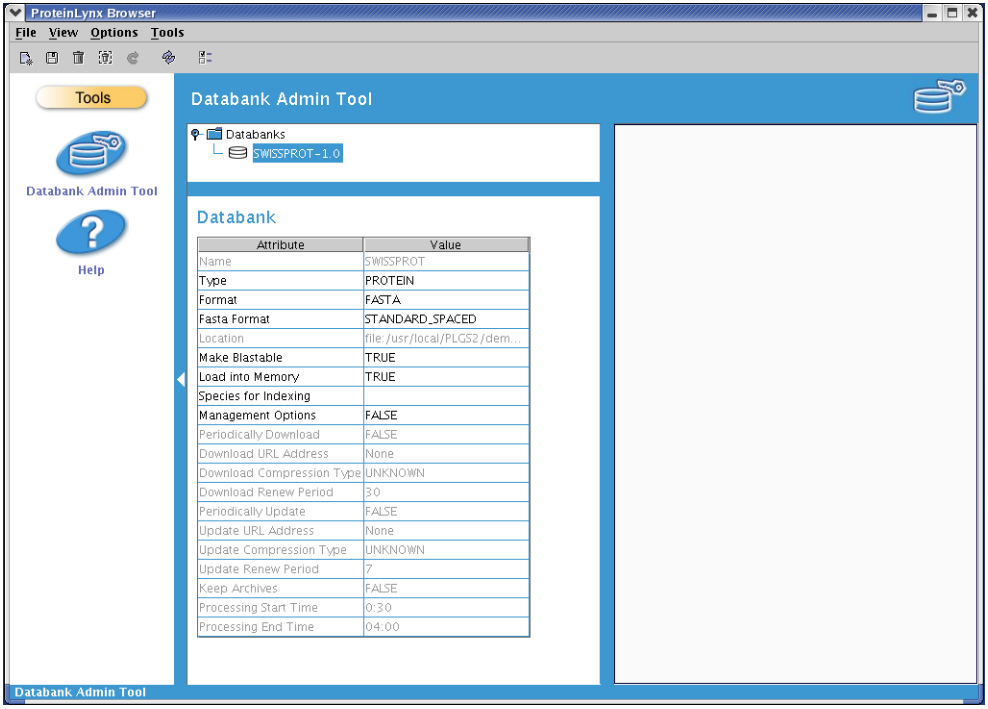
1. Navigate to the `<PLGS install location>/bin` folder.

   #### <PLGS install location>/bin folder:



2. Double-click the ProteinLynxBrowser file to start PLGS.

Linux ProteinLynx browser:



Rule: The Linux ProteinLynx browser supports the Databank Admin and Help tools only.

Starting modules manually and troubleshooting problems

All of the modules are started automatically when you start the ProteinLynx browser on the computer. Nevertheless, you might wish to start the individual modules separately.

To start PLGS modules manually:

1. Navigate to the PLGS installation directory, and then to the bin subdirectory.
2. Start the module by double-clicking in the GUI, or by typing `./<module name>` at the command prompt. At the command prompt, type the following commands:

- `./SearchEngine` to start the search engine.
- `./PLmicrokernel` to start the microkernel.
- `./ProcessorEngine` to start the processor.

If you start the modules automatically, by starting the ProteinLynx browser, log files are generated by the software. These log files can help you to solve operational problems, and will be helpful to Waters if you request technical support.

To view log files:

1. Navigate to the PLGS installation directory, and then to the log subdirectory.
2. Open the log file in a text editor. Two log files are created:
   - Processor.txt for the processor log.
   - SearchEngine.txt for the search engine and microkernel log.

Installing PLGS on UNIX

This section describes the steps required to install, configure, and run PLGS on a non-Linux UNIX computer. PLGS runs on IBM AIX® and Sun Solaris®.

Rule: All UNIX commands are case sensitive.

Before installing PLGS on UNIX

Before installing PLGS on UNIX, you must complete these tasks:

- Back-up the PLGS directories.
- Ensure that you are logged on with root permissions.
- Uninstall previous versions of PLGS.

Backing up the PLGS directory

Before installing PLGS, make a backup copy of the PLGS directory. In a terminal window, type

```
cp -R <source folder> <destination folder>
```

Uninstalling a previous version of PLGS

To uninstall a previous version of PLGS using the command prompt:

1. Go to the old version's _uninst directory by typing

   ```
   cd _uninst
   ```

2. Run the uninstaller by typing

   ```
   /uninstall.bin
   ```

3. Follow the instructions in the Uninstall wizard.

   Tip: After uninstalling PLGS, errors can be reported. This is usually due to the uninstaller not being able to remove the uninstaller resources. This is caused by the user running the uninstaller binary from within the _uninst directory. This means that you will have to remove the _uninst and main PLGS directories manually.

Installing PLGS on UNIX

To install PLGS on UNIX:

1. Insert the PLGS installer CD into the drive.

   Recommendation: Before initializing the installer, copy the installer package from the CD to the local file system.

2. Mount the CD using SMIT, or manually using the **`mount`** command.

   See also: For instructions for mounting the CD, see Appendix D - UNIX Help for Installing PLGS on AIX Platforms.

3. Type the following command in the installer directory:

   `cp PLGS2.2.5_<unix-flavour>.bin <destination>`

   Example: **`cp PLGS2.2.5_SOLARIS_SPARC.bin /usr/local`**

4. Use the `chmod` command to set up permissions on the installer package so that it can be executed:

   `chmod 777 PLGS2.2.5_<unix-flavour>.bin`

   Once the permissions have been set, the installer package is ready to be executed.

5. Type the following command in the directory that the package is in, to execute the installer package:

   `./PLGS2.2.5_SOLARIS_SPARC.bin`

   or

   `./PLGS2.2.5_AIX.bin`

   The installer user interface can take a while to appear. The first welcome screen advises you to ensure that you have uninstalled any previous versions.

   Tip: Occasionally, the installer user interface can appear blank. If this occurs, close down the installer and restart it with the command in step 5.

6. Read and understand the terms of the license agreement. Click Accept in the License Agreement screen, and then click Next. The Destination screen opens.

   Rule: You cannot install PLGS in a directory that has spaces in the name. If you attempt to do so, you will be prompted to enter the path again.

7. In the text field, specify a new or empty directory in which to install the program; the directory should not contain any previous PLGS files. If the directory does not exist, the installer creates the directory automatically.
8. Confirm that your installation details are correct.

   A progress indicator on a splash screen shows the progress of the files being copied to the system.
9. A success message is displayed when the installation is complete.
10. Reboot the machine to ensure that environment variables are setup by the installer. The following SYSTEM environment variables are created:
    - `LIBPATH=<installation path>/lib`
    - `PLGS_HOME=<installation path>`

Configuring PLGS on UNIX

When the installation is complete, to configure PLGS for your specific system you need to:

- Set the number of processors in the mkconfig file (see Databank searching on page 1-25).
- Allocate RAM to the search engine (see Search engine memory allocation on page 1-17).
- Create a TMPDIR environment variable (see TMPDIR environment variable on page 1-18).
- Set a temporary directory for the search engine (see Search engine temporary directory on page 1-18).
- Restore old databanks (see Restoring old databanks on page 1-23).

Search engine memory allocation

When using large databanks with PLGS on a UNIX system, you must alter the amount of RAM allocated to the search engine. You do this by editing the ProteinLynx_SE startup script, which is found in the /bin directory of the installation:

Requirement: Ensure that you have a minimum of 1 GB of RAM before changing the allocation.

To change the memory allocation:

1. Edit the ProteinLynx_SE startup script from:

   ```
   ../jre/bin/java -Xmx256mb
   ```

   to

   ```
   ../jre/bin/java -Xmx1024mb
   ```

2. Save and close the file.

TMPDIR environment variable

Within PLGS is a program called formatdb, which produces the index files necessary for BLAST (Basic Local Alignment Search Tool) searches on a given databank. The program requires an environment variable called TMPDIR to be set to a directory with a large amount of free space. This directory is used as temporary space by formatdb when it is generating the BLAST indices.

To display a list of the environment variables, use the command:

```
set | more
```

If TMPDIR is not displayed in the list, you need to create it. The temporary directory must have read/write permissions.

To create the TMPDIR environment variable:

1. Specify a directory that has 1 GB free space or more:

   ```
   TMPDIR=/tmp
   ```

   where /tmp is the directory with the free space.

2. To enable large databanks to undergo BLAST formatting without any errors, type:

   ```
   export TMPDIR
   ```

Search engine temporary directory

The search engine startup script specifies /tmp as its default temporary directory. This is changed by editing the following entry in the ProteinLynx_SE script:

```
-Duk.co.micromass.searchenginescratch=/tmp
```

Change /tmp to wherever there is a large amount of temporary space available on the system. Typically this could be the same location specified by the TMPDIR variable.

Running PLGS on UNIX

For the AIX version of PLGS there are three components which must be running simultaneously for the system to function. These are the search engine, microkernel, and browser.

Starting the browser automatically starts the other components. Each component can be started manually if required. The browser enables you to add new databanks to the server or view help about the system.

Before running PLGS, ensure that you are logged on with root permissions.

To start the PLGS system:

1. To start PLGS, go to the directory

   <PLGS install location>/bin.

2. Type **./ProteinLynxBrowser** to start the browser.

Restriction: Only the Databank Admin Tool and the online Help are available in the UNIX PLGS browser. If so configured, processing and searching can be run on a UNIX machine from a remote Windows PLGS browser.

Starting modules manually and troubleshooting problems

All of the modules are started automatically when you start the ProteinLynx browser on the computer. Nevertheless, you might wish to start the individual modules separately, however.

Before running PLGS, ensure that you are logged on with root permissions.

To start modules manually:

1. Go to the directory

   <PLGS install location>/bin

2. Start the modules by typing the following at the command prompt:
   - **./SearchEngine** to start the search engine.
   - **./PLmicrokernel** to start the microkernel.

- **`./ProcessorEngine`** to start the processor.

If you start the modules automatically, by starting the ProteinLynx browser, log files are generated by the software. These log files can help you to solve operational problems, and will be helpful to Waters if you request technical support.

To view log files:

1. Go to the directory

   `<PLGS install location>/log`

2. Open the log file in a text editor.
   - Processor.txt for the processor log.
   - SearchEngine.txt for the search engine and microkernel log.

Installation troubleshooting on UNIX

The following sections detail possible causes and solutions regarding installation problems on UNIX.

Installer startup problems

The installer package can fail to start if there is insufficient temporary space in its current directory. To remedy this, either run the installer package from another directory or specify the following command line arguments when running the installer:

`/PLGS2.2.5_<unix-flavour>.bin -is:tempdir /tmp` `(where /tmp is a directory with lots of free space)`

If this does not solve the problem, check that the installer package has full permissions by using:

`chmod 777 PLGS2.2.5_<unix-flavour>.bin`

If the problem persists, the file could have been corrupted while being copied from the CD.

Microkernel failures

If the microkernel fails to start, check the following:

- Check that your system is enabled for 64 bit operation; this can be done from the ‘smit’ application. If the system is not enabled for 64 bit

operation, it might display messages about incorrect libraries when starting the microkernel.

- Check that the permissions levels on the PLmicrokernel file are sufficient. If not, change the permissions by typing the following command in the file:

  ```
  chmod 777 PLmicrokernel
  ```

- Check that the number of processors specified in the config/micro/mkconfig file are appropriate (see Setting the number of processors on page 1-24).
- Ensure you are logged on as root.
- Ensure user root has read/write and execute permissions on the databanks and their associated files.

Recommendation: Index files that are created by databanks should be in the same directory as the databanks.

Search engine failures

If error traces are seen in the console window or log file of the search engine, ensure that you have selected the correct format for all databanks added to the server (see Databank attributes on page 13-4).

Large databank (>2 GB) problems

If you experience problems when searching or adding large databanks, check the following:

- Check that large file support is enabled on the temporary space (the directory is specified in the search engine startup script).
- Check that large file support is enabled on the directory that contains the databanks.
- Check that the search engine has 2 GB of RAM allocated to it. See Search engine memory allocation on page 1-17 for details.

Databank and BLAST searching problems

If problems occur with databank or BLAST searching, try carrying out the following operations:

- Remove user account file-size restrictions.

- Increase the amount of space allocated to a particular mount point.
- Enable LARGE_FILE support for the mount point. This can be done using the system administration tool.
- Remove limits on memory allocation for a user account. This can also be done using the system administration tool.

If you are unsure how to perform these tasks, check with your UNIX administrator.

Restoring old databanks

When performing a new installation, any databanks added to previous versions are not available from the new PLGS version. The databanks must be restored using the Databank Admin tool. This tool allows you to specify the format of the databank (usually FASTA), and the sub-format of the databank, (such as NCBI_EXPASY_GENERAL).

Caution: If an incorrect databank format is specified the databank will not be added correctly, which can subsequently cause problems with PLGS.

To determine the type of databank, view the first line of the databank in a terminal window by using:

```
more <databank name>
```

For information on the various formats available, see FASTA flat file format on page E-9.

Setting the number of processors

If the computer on which you are installing ProteinLynx Global SERVER has more than one processor, you can take advantage of the additional power with PLGS.

Tip: If your computer only has one processor, or if you wish PLGS to only use one processor, you do not need to make any changes.

The number of processors used can be individually set for three different circumstances:

- DDA data processing
- Expression data processing
- Databank searching

Caution: Never set the number of processors to a value greater than the number of processors on your system.

DDA data processing

Recommendation: Make a copy of the file before editing, as making changes other than those explicitly outlined below could prevent PLGS from operating properly.

To set the number of processors for DDA processing:

1. Navigate to the `lib` directory, underneath the PLGS installation directory.
2. Open the process.cfg file. If it does not exist, create a text file called process.cfg, and then open it.
3. Add the following lines to the file:

   ```
   [MULTITASKING]
   Number of Processors=<number>
   ```

   Where `<number>` is the number of processors you want DDA processing to utilize.
4. Save the file.

Expression data processing

Recommendation: Make a copy of the file before editing, as making changes other than those explicitly outlined below could prevent PLGS from operating properly.

To set the number of processors for Expression processing:

1. Navigate to the `lib` directory, underneath the PLGS installation directory.
2. Open the process.cfg file. If it does not exist, create a text file called process.cfg, and then open it.
3. Add the following lines to the file:

   ```
   [EKL Processing]
   Number of Processors=<number>
   ```

   Where `<number>` is the number of processors you want Expression data processing to utilize.
4. Save the file.

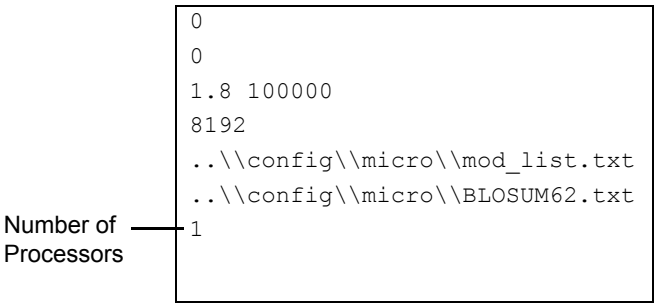
Databank searching

Recommendation: Make a copy of the file before editing, as making changes other than those explicate outlined below could prevent PLGS from operating properly.

To set the number of processors for databank searching:

1. Navigate to the `config\micro` directory, underneath the PLGS installation directory.
2. Open the mkconfig file.

The file contains the following lines:



3. On the seventh line of the file, type the number of processors you want databank searching to utilize.
4. Save the file.

2 Setting up ProteinLynx Global SERVER

You can set up the ProteinLynx Global SERVER browser for the way you want to work; this includes:

- Adding and removing tools from the Tool tray.
- Identifying search engines, processors and instruments that are to be used to process data.
- Specifying Uniform Resource Locators (URLs) for Web sites that can be referenced within the application.
- Setting the colors for the display of the microtitre and target plates.
- Setting the style and display for printing results.
- Specifying the location of modules used in automated processes, and altering the behavior of these modules.
- Specifying additional formats in which spectra can be saved after processing.
- Altering the modules (PlugIns) that handle archiving and retrieval of ProteinLynx project data.

Contents:

Topic	Page
ProteinLynx browser	2-2
Changing preferences	2-5
Setting Automation Setup parameters	2-18

ProteinLynx browser

The user interface for PLGS is the ProteinLynx browser, which provides access to various PlugIn tools in the ProteinLynx suite (see Figure titled “ProteinLynx browser:” on page 2-3).

The ProteinLynx browser enables you to:

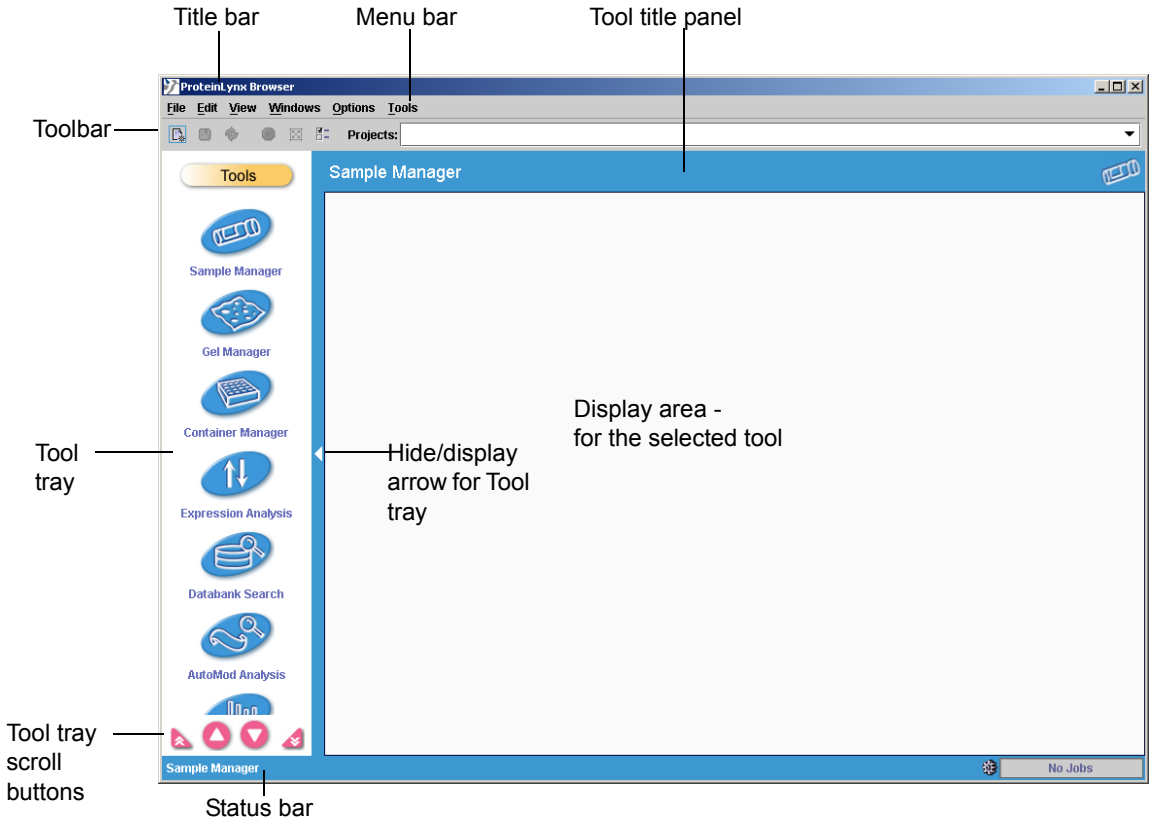
- View and edit global preferences.
- View and edit automation set-up parameters.
- Change between tools.
- Manage the desktop, which is shared by most of the tools.

The content of the toolbar and menus varies depending on which tool is selected.

The Preferences button , is the only button common to all toolbars. The following commands are common throughout the software from the Menu Bar:

- File > Exit
- Options > Preferences (see Changing preferences on page 2-5)
- Options > Automation Setup (see Setting Automation Setup parameters on page 2-18)
- Tools > Add/Remove Tools (see Adding and removing tools on page 2-4)

ProteinLynx browser:



Tool tray

The tool tray provides links to all the available tools. Use the buttons at the bottom of the Tool tray to navigate through the list of tools (see Scroll buttons for the tool tray: on page 2-4).

To hide or display the tool tray, click the arrow or on the splitter bar between the tool tray and the Display Area.

Note: Some tools could have been removed from the list using the Add/Remove Tools menu (see Adding and removing tools on page 2-4). Therefore, there might be fewer tools displayed than those shown in ProteinLynx browser: on page 2-3.

The following table details the scroll buttons for the tool tray.

Scroll buttons for the tool tray:

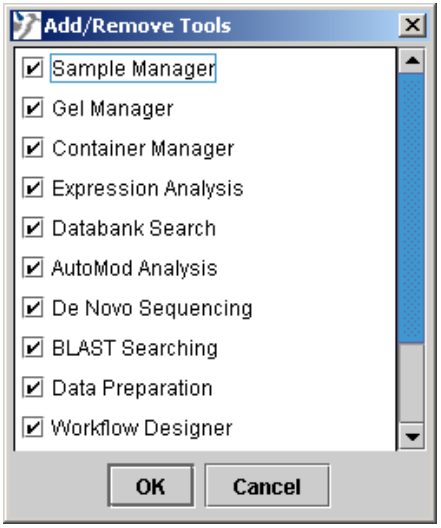
Button	Action
	Displays the top section of the tool tray.
	Scrolls up the list of tools.
	Scrolls down the list of tools.
	Displays the bottom section of the tool tray.

Adding and removing tools

To customize the list of tools shown in the Tools menu and tool tray:

1. Click Tools > Add/Remove Tools.

Add/Remove Tools dialog box:



2. Select or clear the check box for each tool to include or exclude the tool in the Tools menu and tool tray.

Changing preferences

The ProteinLynx Browser Preferences dialog box enables you to change preferences for the search engine, processors, instrument type, bookmarks, plate colors and printing.

To open the ProteinLynx Browser Preferences dialog box, either:

- On the toolbar, click , or
- Click Options > Preferences.

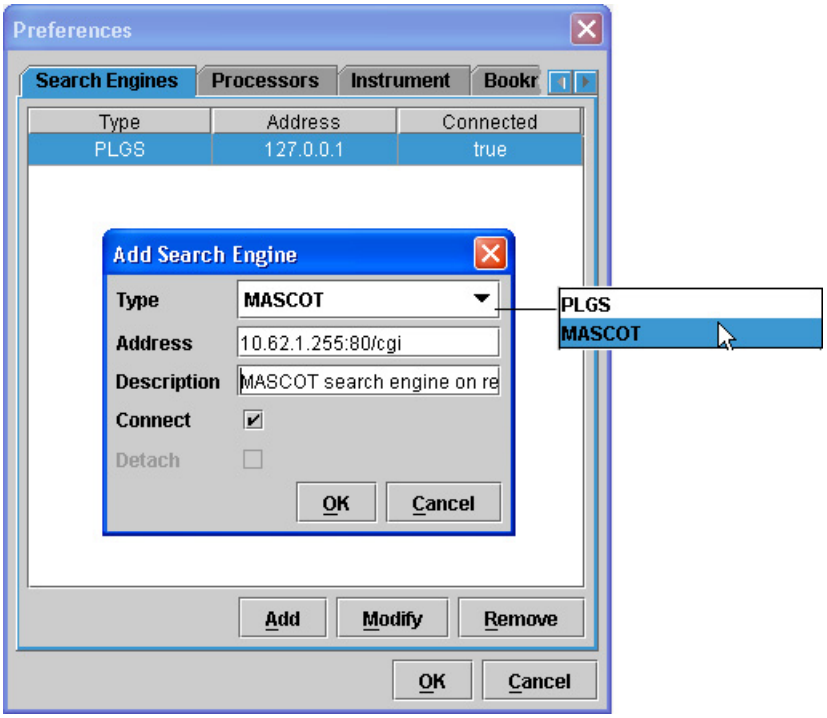
The dialog box has a number of tabs:

- Search Engine (see Search Engine tab on page 2-5) – enables you to add, remove, or select a search engine.
- Processors (see Processors tab on page 2-8) – enables you to add, remove, or select multiple processors.
- Instrument (see Instrument tab on page 2-10) – enables you to change the current type of instrument.
- Bookmarks (see Bookmarks tab on page 2-11) – enables you to specify bookmarks that can be accessed from other parts of the system.
- Colours (see Colours tab on page 2-12) – enables you to view and edit the plate colors.
- Printing (see Printing tab on page 2-16) – enables you to specify settings for printing the project or workflow data.

Search Engine tab

Use this tab to add, remove, or select a search engine.

Preferences dialog box, Search Engine tab:



ProteinLynx browser can submit searches to PLGS or MASCOT (version 2.0 and later) search engines, running either on the local PC (IP address 127.0.0.1) or on remote servers.

Adding a search engine

You can add one search engine of each type, PLGS or MASCOT.

To add a search engine:

1. Click Add.
2. Click the type of search engine: PLGS or MASCOT.

3. Type or paste the IP address of the computer, on which the search engine is running, into the Address text box.

   To connect to a PLGS server, you only have to type the IP address. However, to connect to a MASCOT server, you must type the IP address, port number and the path to the CGI (Common Gateway Interface) directory. For example:

   10.62.1.255:80/cgi

   Tip: Port 80 and 8080 are commonly used for internet applications, including Mascot. If port 80 or 8080 are not correct, please consult your Mascot server administrator.

   The CGI directory contains the program that executes the databank search. The default location of this directory is <IP address>/mascot/cgi. However, it is recommended that you consult your Mascot server administrator to check the location of the directory.

4. Type a description of the search engine in the Description text box.
5. To connect immediately, select Connect.
6. If you want the search engine to keep running when the ProteinLynx browser is closed, select Detach.
7. Click OK.

Modifying a search engine

You can modify the type of search engine, IP address, description, and the connection details of a search engine.

To modify a search engine:

1. Double-click the search engine in the list.

   Alternative: Click the search engine, and then click Modify.

2. The Modify Search Engine dialog box opens, which has the same fields as the Add Search Engine dialog box.
3. Modify the details as required.
4. Click OK.

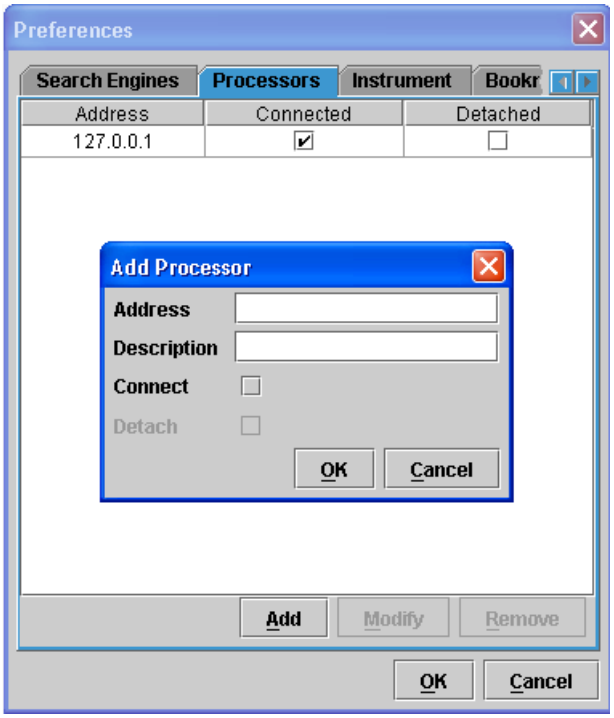
Removing a search engine

To remove a search engine, click the search engine, and then click Remove.

Processors tab

Use this tab to add, modify, or remove local or remote processors. The browser can process raw data on the host machine or on remote processors. However, the Processor module must be running on the same computer as the raw data. The details of any remote processor must be entered in the Processors page on the host machine.

Preferences dialog box, Processors tab:



Adding a processor

You can add local or remote processors.

To add a processor:

1. Click Add.
2. In the Address text box, type or paste the IP address of the computer on which the processor is running.
3. In the Description text box, type a description of the processor.

   Example: “Remote processor on UNIX box 2”.
4. To connect immediately, select Connect.
5. If you want the processor to keep running when the ProteinLynx browser is closed, select Detach.
6. Click OK.

Modifying a processor

You can modify the IP address, description, and the connection details of a processor.

To modify a processor:

1. Double-click the processor in the list.

   Alternative: Click the processor, and then click Modify.
2. The Modify Processor dialog box opens, which has the same fields as the Add Processor dialog box.
3. Modify the details as required.
4. Click OK.

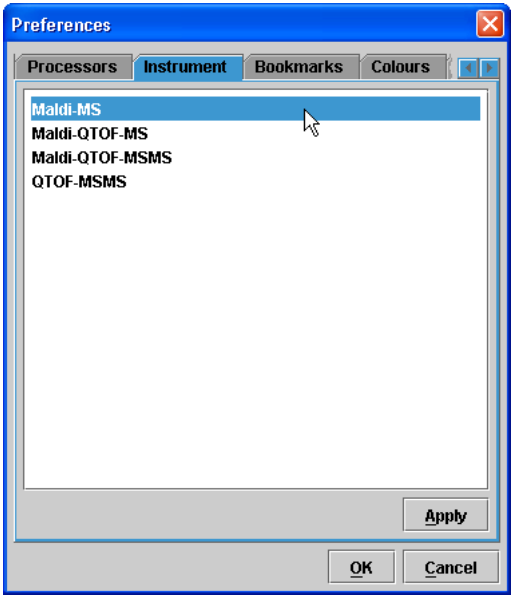
Removing a processor

To remove a processor, click the processor, and then click Remove.

Instrument tab

Use the Instrument tab to change the current type of instrument. This specifies the instrument from which raw data is acquired, and can affect various default values: for example, the default processing parameters used for spectrum data will depend on the instrument type.

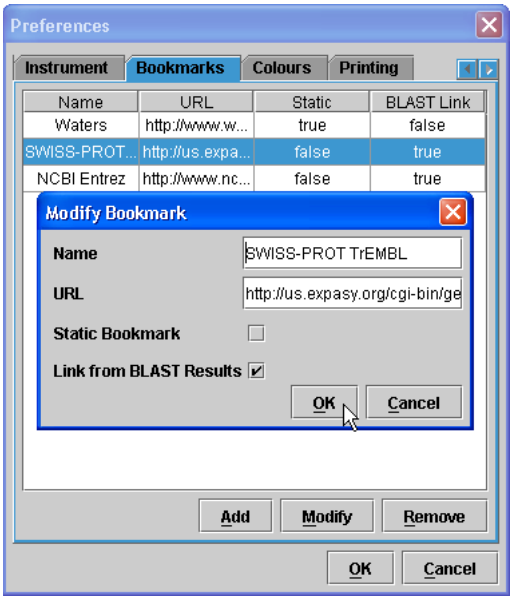
Preferences dialog box, Instrument tab:



Bookmarks tab

Use the Bookmarks tab to specify URLs for access elsewhere in the system.

Preferences dialog box, Bookmarks tab:



Adding a bookmark

You can add static or dynamic bookmarks to the list.

To add a bookmark:

1. Click Add to open the Add Bookmark dialog box.
2. In the dialog box, type the name of the bookmark and the URL.
3. Select the Static Bookmark check box if the bookmark is static (always the same), or clear the Static Bookmark check box if the bookmark is dynamic.

   A dynamic bookmark is not a valid URL until it is combined with a unique identifier. For example, to form a valid URL, the SWISS-PROT TrEMBL link that is supplied with ProteinLynx browser requires the addition of an accession number. This URL then provides a link to the SWISS-PROT TrEMBL databank entry for the specified accession number.

4. Select or clear the Link from BLAST Results check box.

   If selected, hyperlinks to the external database can be formed from accession numbers returned from BLAST (Basic Local Alignment Search Tool) searches.

5. Click OK to save the changes.

Modifying a bookmark

You can modify the name, URL, static bookmark status, and BLAST results link status of a bookmark.

To modify a bookmark:

1. Double-click the bookmark in the list.

   Alternative: Click the bookmark, and then click Modify.

2. The Modify Bookmarks dialog box opens, which has the same fields as the Add Bookmark dialog box.
3. Modify the details as required.
4. Click OK.

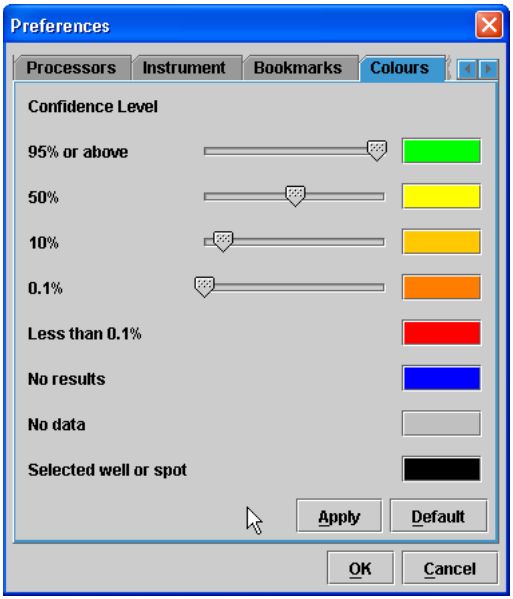
Removing a bookmark

To remove a bookmark, click a bookmark, and then click Remove.

Colours tab

Use the Colours tab to view and edit the well or spot colors that are shown in the target plate graphic in the Container Manager display (see Creating a new vial, microtitre or target plate on page 5-9). The colors show the status of a microtitre plate well or target plate spot and, when appropriate, the confidence level of the top scoring hit.

Preferences dialog box, Colours tab:



The confidence levels and colors shown are the defaults.

Default plate color descriptions:

Well or Spot State	Confidence Level	Color
High score	95% or above	Green
Medium score	50%	Yellow
Medium-low score	10%	Light orange
Low score	0.1%	Orange
Very low score	Less than 0.1%	Red
No results		Blue
No data		Gray
Selected well or spot		Black

Setting confidence levels and colors

You can adjust the confidence levels of results that trigger the display of the colors in the wells or spots.

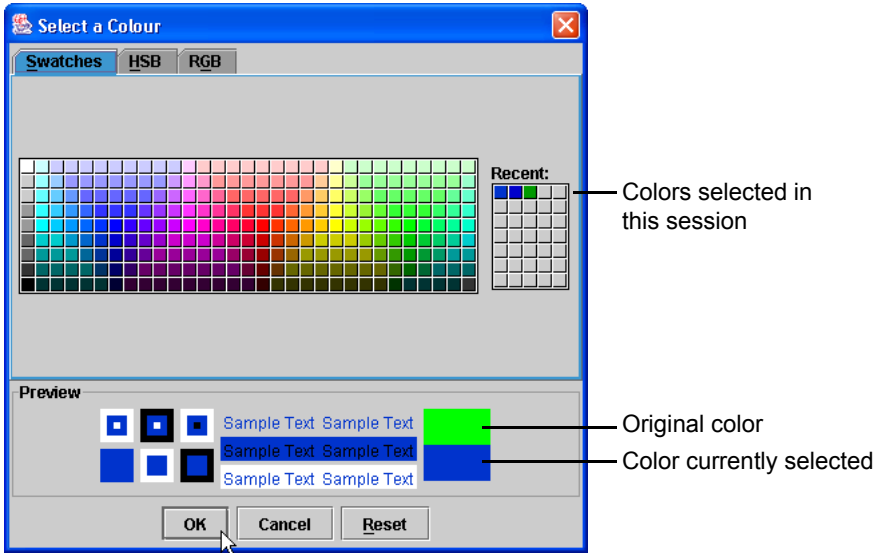
To set the confidence levels and colors:

1. Use the slider bars to adjust confidence levels.
2. To change a color associated with a confidence level, click the color.

   The Select a Colour dialog box opens. This dialog box has three tabbed pages, any of which can be used to select the color:

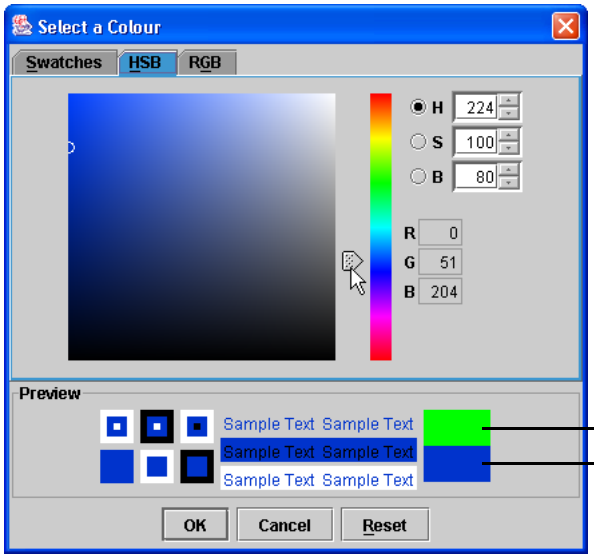
   - Swatches — Enables you to select from a panel of predefined colors.
   - HSB — Enables you to select a color using the Hue-Saturation-Brightness (HSB) color model.
   - RGB — Enables you to select a color using the Red-Green-Blue (RGB) color model.

Select a Colour dialog box- Swatches tab:

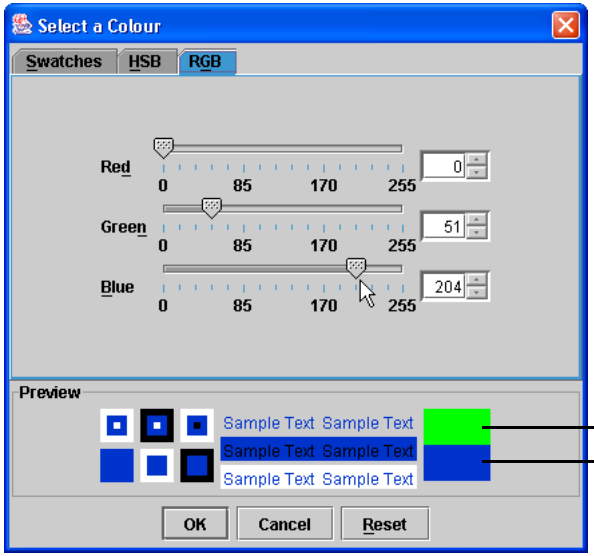


The Recent: section shows the colors that you have selected in this session.

Select a Colour dialog box - HSB tab:



Select a Colour dialog box- RGB tab:



For each page:

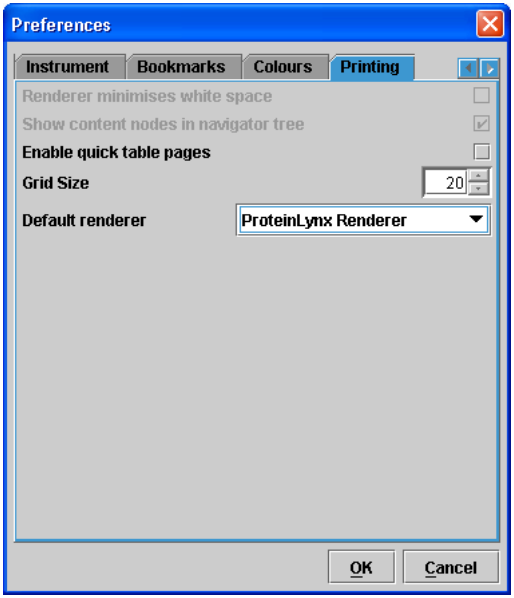
- The Preview pane shows how the color selected will look in different situations. The top half of the block to the right shows the original color when this dialog box was opened; the bottom half shows the color currently selected.
- The Reset button resets the color to the original.

3. To set the color you have selected, click OK.

Printing tab

Use the Printing tab to view and edit the printing preferences.

Preferences dialog box, Printing tab:



Restriction: The dimmed options are not available in this version of PLGS.

To edit the printing preferences:

1. To be able to add tabular as well as graphical data to a print template, select the 'Enable quick table pages' option.

   This enables the option Tabular Data in the Template Type dialog box when creating new templates (see Creating print templates on page 11-13). Selecting this also enables you to add tables to Results nodes table pages in the Print Tool navigator tree when creating new templates (see Adding content to the results nodes on page 11-15).

2. To change the size of the grid in the page editor view, type or scroll to a number in the Grid Size option. See Customizing print templates on page 11-19 for details of how to use the grid.

3. To change the print renderer for different applications, select from the drop-down list.

   This changes the renderer for any new templates that you create. However, existing templates will use the renderer that was originally applied to that template.

Setting Automation Setup parameters

The configurable parameters in the ProteinLynx Browser Automation Setup dialog box are used by modules that handle automated data acquisition, processing, and searching.

To open the ProteinLynx Browser Automation Setup dialog box from the menu bar, click Options > Automation Setup.

The dialog box has three tabs:

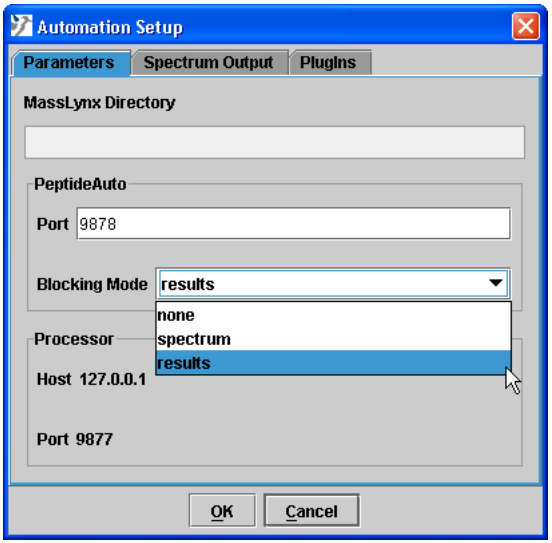
- Parameters (see Parameters tab on page 2-18) – enables you to specify the location of modules used in automated processes, and alter the behavior of these modules.
- Spectrum Output (see Spectrum Output tab on page 2-20) – enables you to specify additional formats in which spectra can be saved after processing.
- Plugins (see PlugIns tab on page 2-23) – enables you to alter the modules (Plugins) that handle the archiving and retrieval of ProteinLynx project data.

Parameters tab

A key feature of the ProteinLynx system is its ability to fully automate the acquisition, processing, and searching of data. The Parameters tab enables you to specify the location of modules used in automated processes, and alter the behavior of these modules.

To update the settings, click OK.

Automation Setup dialog box, Parameters tab:



You can set the following parameters.

Parameters tab parameters:

Parameter	Description
MassLynx Directory	Type the pathname of the directory in which MassLynx is installed on the local PC.
PeptideAuto - Port	The port enables the application to interface with other modules. Type the port number used by the PeptideAuto module. PeptideAuto handles submission of data for processing, and workflows for searching, from MassLynx. **Recommended:** Use the default port number.

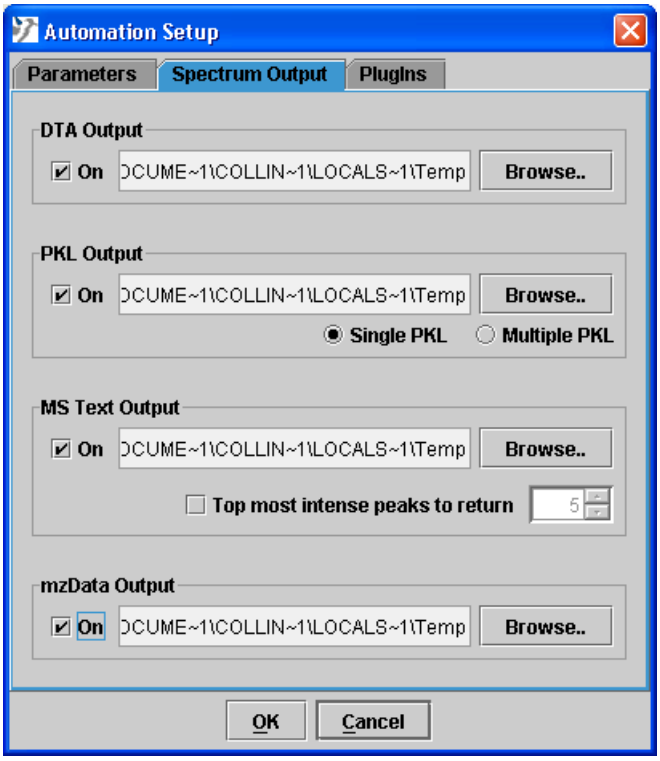
Parameters tab parameters: (Continued)

Parameter	Description
PeptideAuto - Blocking Mode	The blocking mode parameter describes the data acquisition behavior of MassLynx. The following blocking modes are available: • none - MassLynx will continue to acquire data while previously acquired data is being processed or used for searches. • spectrum - MassLynx data acquisition will be blocked until any previous data has been processed (although data can still be acquired while previous data is being used for searches). • results - MassLynx data acquisition will be blocked until any previous data has been processed, and until any searches using the previously acquired data are complete. **Recommendation:** The preferred option depends upon the hardware configuration. For example, if searching is being performed on a remote server, do not block on results, as the acquisition PC would be free to continue acquisition during the data search step.
Processor - Host	Type the IP address of the computer on which the processor is running. The processor module handles processing of raw data to produce mass spectra. **Tip:** This information is for the local processor. Use the Preferences dialog box (see Processors tab on page 2-8) to specify remote processors.
Processor - Port	Type the port number used by the processor module.

Spectrum Output tab

The Spectrum Output tab enables you to specify additional formats in which spectra can be saved after processing. Spectra are automatically saved in ProteinLynx XML format.

Automation Setup dialog box, Spectrum Output tab:



You can set the following parameters.

Spectrum Output tab parameters:

Parameter	Description
DTA Output	DTA format is a Waters file format for storing MS/MS spectra. The first line of a DTA format file contains the singly protonated peptide mass (MH^+) and the peptide charge state as a pair of space separated values. Subsequent lines contain space separated pairs of fragment ion m/z and intensity values. In a DTA file, the precursor peptide mass is an MH^+ value independent of the charge state. In Mascot generic format, the precursor peptide mass is an observed m/z value, from which M_r or MH_n^{n+} is calculated using the prevailing charge state. Include at least one blank line between each MS/MS dataset. For more details, see www.matrixscience.com.
PKL Output	PKL format is a Waters file format for storing MS/MS spectra. The PKL format is similar to the DTA file format, but supports multiple MS/MS datasets in a single file. The first line of a PKL dataset contains the observed m/z, intensity, and charge state of the precursor peptide as a triplet of space separated values. Subsequent lines contain space separated pairs of fragment ion m/z and intensity values. Multiple MS/MS datasets are delimited by at least one blank line.
MS Text Output	MS Text format is a plain text file, listing mass-intensity pairs, suitable for storing an MS spectrum. If this is selected, the Top most intense peaks to return check box is enabled.

Spectrum Output tab parameters: (Continued)

Parameter	Description
mzData Output	The mzData format contains information similar to that in the PKL format, but in an open source XML format that is supported by various other scientific software providers. **See also:** The Proteomics Standards Initiative's website at http://psidev.sourceforge.net/ms/ .

To add a format:

1. Select the check box next to the name of the format.
2. Click Browse.., and then select a folder where the spectra output is to be saved.

   If the MS Text Output format is specified, the Top most intense peaks to return check box is enabled. Selecting the check box enables you to specify the maximum number of peaks written to the MS Text Output file. If the check box is not selected, the mass-intensity pairs of all peaks will be written to the MS Text Output file.

PlugIns tab

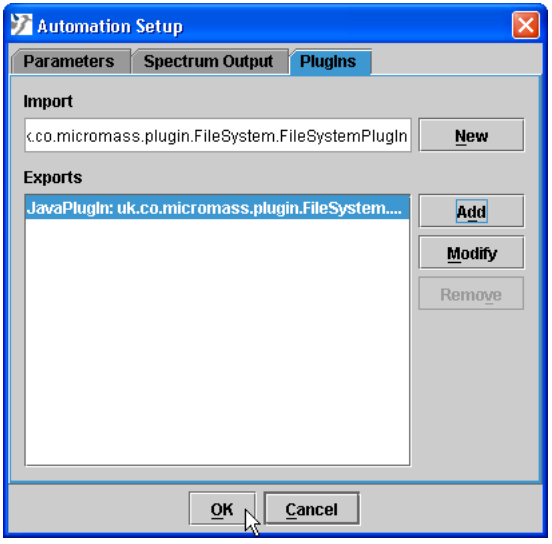
In PLGS, all of the data representing a project (gels, containers, spectra, queries, results, and so on) is archived through a supplied PlugIn, which saves these projects locally in XML format. However, it is possible to replace this plugin or add additional third party plugins to handle the project XML in a different manner; to parse and write it into a format more suitable for your needs.

- Import – To save data from other sources and formats into a PLGS project.
- Export – To retrieve data from PLGS projects and export the data to other formats.

An example of a PlugIn is the FileSystemPlugIn, which is supplied with PLGS. This PlugIn is used to import data from other sources into the standard PLGS file structure. This PlugIn also exports data from the standard PLGS file structure into other formats.

For more details of the implementation and use of PlugIns, see Appendix C - Implementing a plugin for ProteinLynx Global SERVER.

Automation Setup dialog box, PlugIns tab:



Replacing the Import PlugIn or adding an Export PlugIn

You can replace the supplied Import PlugIn, but you cannot modify it or add more Import PlugIns. However, you can modify the supplied Export PlugIn and add new Export PlugIns.

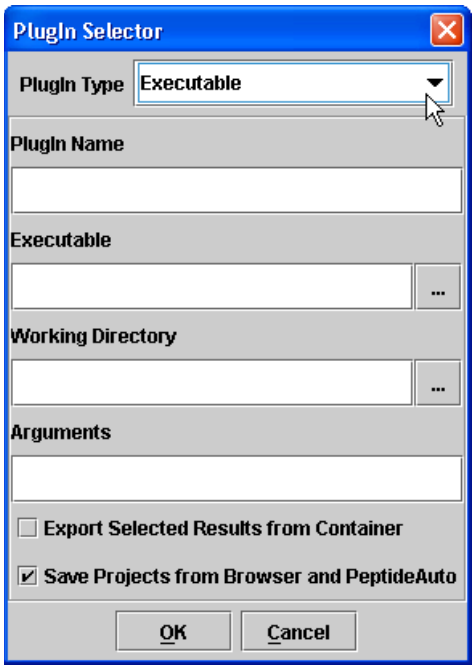
The dialog boxes are the same for replacing the Import PlugIn and adding Export PlugIns.

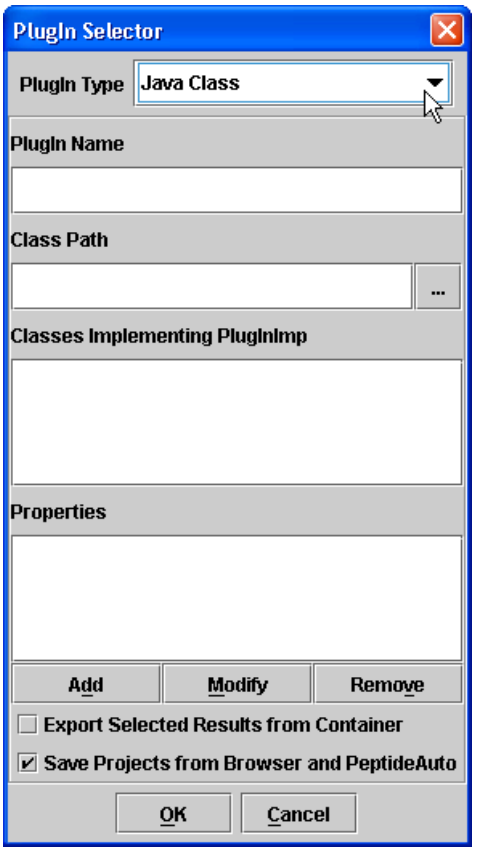
To replace the Import PlugIn or add an Export PlugIn:

1. Click New to replace the Import PlugIn, or click Add to add another Export PlugIn.

   You can select from two types of PlugIn: Executable or Java Class, which have different attributes.

PlugIn Selector dialog boxes - Executable and Java Class PlugIn types:





2. Add the details to the attribute fields for the Executable or Java Class PlugIn.

Attributes - Executable PlugIn:

Attribute	Description
PlugIn Name	Optional — Required only if you want to export results from a container directly to this PlugIn, bypassing the FileSystemPlugIn and any other third-party PlugIns.

Attributes - Executable PlugIn:

Attribute	Description
Executable	Click to browse for the location of the executable, or type the full path to the executable.
Working Directory	Click to browse for the location of the directory to which you want the PlugIn to write its files, or type the full path to the directory.
Arguments	Type the list of command line arguments required by the PlugIn.
Export Selected Results from Container	Select this to export selected results from a container directly to the PlugIn. **Default:** Cleared.
Save Projects from Browser and PeptideAuto	Select this to execute the PlugIn whenever projects are updated by the browser or PeptideAuto. **Default:** Selected.

Attributes - Java Class PlugIn:

Attribute	Description
PlugIn Name	Optional — Required only if you want to export results from a container directly to this PlugIn, bypassing the FileSystemPlugIn and any other third-party PlugIns.
Class Path	Click to browse for the location of the *.jar file or class, or type the full path to the *.jar file or class.

Attributes - Java Class PlugIn: (Continued)

Attribute	Description
Classes Implementing PlugInImp	When the plugin's jar or class file has been declared in the Class Path field the list of classes found in the plugin that implement the interface PlugInImp are displayed. This is for your information only and is there only to confirm that the plugin does implement this class.
Properties	You can add, remove or modify any properties required by the PlugIn, for example, the working directory of the PlugIn. To add or modify a property, click Add or Modify. Type the values in the Add/Modify dialog box that opens. To remove a property, select the property, and then click Remove.
Export Selected Results from Container	Select this to export selected results from a container directly to the PlugIn. **Default:** Cleared.
Save Projects from Browser and PeptideAuto	Select this to execute the PlugIn whenever projects are updated by the browser or PeptideAuto. **Default:** Selected.

3. In the PlugIn Selector dialog box, click OK.

   Result: For an Import PlugIn, the new PlugIn replaces the previous PlugIn. For an Export PlugIn, the new PlugIn is added to the list.

4. On the PlugIns tab, click OK.

   Requirement: For the PlugIn to work, the ProteinLynx Browser must be restarted.

Modifying an Export PlugIn

You can modify the details of any Export PlugIn, including the supplied PlugIn.

To modify an Export PlugIn:

1. On the PlugIns tab, select the PlugIn from the list.
2. Click Modify. The PlugIn Selector dialog box opens (Figure titled "PlugIn Selector dialog boxes - Executable and Java Class PlugIn types:" on page 2-25), which contains the details of the PlugIn.
3. Modify the details as required, and then click OK.
4. On the PlugIns tab, click OK.

   Requirement: For the PlugIn changes to take effect, the ProteinLynx browser must be restarted.

Removing an Export PlugIn

Rule: You can only remove an Export PlugIn when there is more than one in the list.

To remove an Export PlugIn:

1. In the PlugIns page, select the PlugIn from the list, and then click Remove. The PlugIn is removed from the list.
2. Click OK.

   Requirement: For the PlugIn changes to take effect, the ProteinLynx Browser must be restarted.

3 Creating, importing, and managing projects

You organize your work in ProteinLynx Global SERVER using projects. Each project contains a collection of related settings, files, and data that represent an area of work.

Many of the tools you work with in PLGS create and manage settings and templates that can be applied across projects. These tools do not require a project to be created or opened.

Sample Manager, Gel Manager, Container Manager, and Expression Analysis require that a project is created or opened before they can be used.

Contents:

Topic	Page
Creating a new project	3-2
Importing and exporting projects	3-3
Opening and updating projects	3-5
Closing and deleting projects	3-6

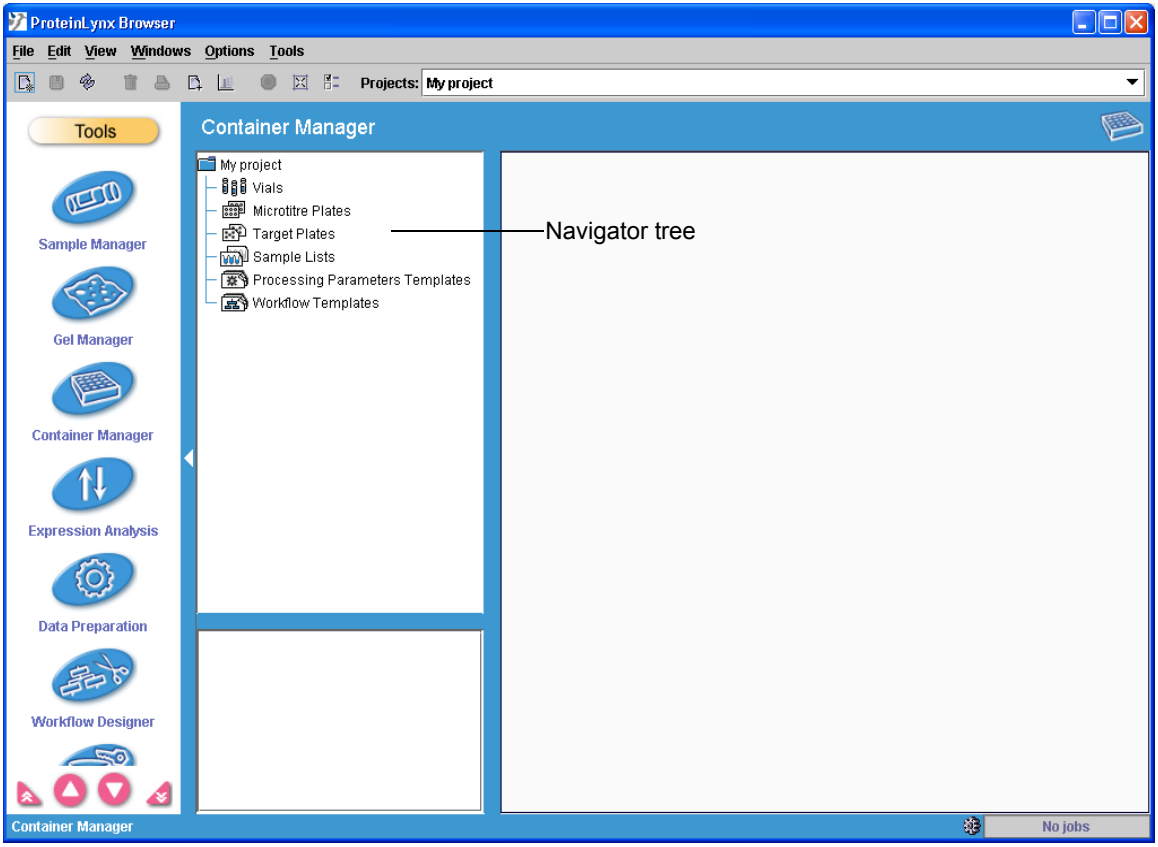
Creating a new project

To create a project:

1. In the tool tray, click the icon for one of the tools that requires a project: Sample Manager, Gel Manager, Container Manager, or Expression Analysis.
2. Click the Create new project button on the toolbar.
3. Type a name for the project.
4. Click OK.

Result: The Container Manager window looks similar to the following illustration.

Container Manager with new project:



Importing and exporting projects

To import a project:

1. In the tool tray, click the icon for one of the tools that requires a project: Sample Manager, Gel Manager, Container Manager, or Expression Analysis.
2. Click File > Import Project.
3. Click the Files of Type drop-down list, and then click the type of project file you want to import.
   - PDQuest XML – Sample list XML file generated from PDQuest software. Importing this file type imports any gel, container, and sample tracking information specified in the XML.
   - Progenesis XML – Experiment XML file generated from Progenesis Discovery software. Importing this file type imports any project and gel information specified in the XML.
   - XML file – The ProteinLynx Global SERVER project XML file. Using this import option allows you to explicitly specify project and project member ids. The XML is validated against the Protein Lynx Global Server XML schema.

     Caution: This option will not import data or results. It should only be used to import a skeleton project that includes sample and container information.
   - ZIP file – A ProteinLynx Global SERVER zipped project created by exporting a project from PLGS.
4. Click Open.

Result: The project is imported into PLGS, and then opened. Depending on the size of the project imported, the process can take some time. The status bar in the bottom right of the browser indicates that the import is in progress.

To export a project:

1. Click File > Export Project.
2. Navigate to the directory in which you want to save the exported project, and type a name for the file.
3. Click Save.

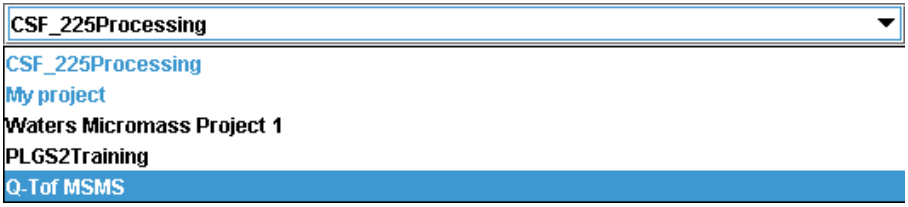
Result: The project is exported as a compressed .zip file, which can then be imported into another PLGS installation.

Opening and updating projects

To open a project:

1. In the tool tray, click the icon for one of the tools that requires a project: Sample Manager, Gel Manager, Container Manager, or Expression Analysis.
2. Click the Projects box, in the PLGS toolbar, to display the projects list.

   ### Example projects list:

3. Click a project to display it in the browser.
   - Project names in black text are available, but not currently open.
   - Project names in blue text are currently open.
   - Project names in gray text are unavailable: they cannot be opened. Projects might be unavailable because they are currently being saved or deleted.

Updating projects

When MassLynx is used to acquire data based on information exported from ProteinLynx Global SERVER, PLGS projects can be updated to reflect the most recent information available. Updating projects is not usually necessary at other times.

To update a project:

In the ProteinLynx browser, click File > Update.

Closing and deleting projects

To close a project:

1. In the tool tray, click the icon for one of the tools that requires a project: Sample Manager, Gel Manager, Container Manager, or Expression Analysis.
2. If the project is not currently displayed, switch to the project you wish to close (see To open a project: on page 3-5 for details).
3. Click File > Close.

Result: The selected project is closed, releasing any resources it is using and closing any associated windows.

Rule: If changes have been made since the project was last saved, you can save the project before it is closed.

To delete a project:

1. In the tool tray, click the icon for one of the tools that requires a project: Sample Manager, Gel Manager, Container Manager, or Expression Analysis.
2. If the project is not currently displayed, switch to the project you wish to delete (see To open a project: on page 3-5 for details).
3. Click the name of the current project in the navigator tree.
4. Click Edit > Delete.
5. If you are sure you want to delete the project, click Yes.

Result: The project is deleted, and is no longer available in the ProteinLynx browser. Processed data is deleted, but the original raw data is not.

4 Annotating and tracking samples with Sample Manager

Sample Manager enables the full annotation and tracking of all the samples used in a ProteinLynx project.

Contents:

Topic	Page
Getting started with Sample Manager	4-2
Sample editor	4-3

Getting started with Sample Manager

The Sample Manager enables you to fully annotate all the samples used in a ProteinLynx project. Individual samples can be named and associated with hyperlinks, allowing clear sample tracking throughout the whole ProteinLynx system. Also, individual samples can be mixed to produce processed samples, which include full details of their origin.

When you set a sample in Container Manager (see What is Container Manager? on page 5-2), you choose from the samples that you added to Sample Manager. The samples specified and configured in Sample Manager are also those identified for use in Expression experiments.

To open the Sample Manager, click the Sample Manager icon on the tool tray.

Adding a sample

To add a sample to a project:

1. In the navigator tree click, and then right-click Original Samples.
2. Click Add New Sample.
3. You are asked whether you want to add the new sample to a new vial. Click Yes or No.

   Rationale: Whether you choose Yes or No, a new sample is produced, its details are displayed, and it is added to the navigator tree. Clicking Yes also produces a new vial in the Container Manager to which the new sample is added.

Deleting a sample

To delete a sample:

1. Click a sample in the navigator tree.
2. Click Delete on the toolbar.

Restriction: You can only delete samples that are not being used anywhere else on the system.

Sample editor

To modify or view the information associated with a sample, highlight the sample name in the navigator tree. The Sample Editor is displayed.

Sample Manager - sample editor:

To add or modify an attribute:

1. Click the attribute in the panel.
2. Enter the value at the bottom of the panel.

Restriction: You cannot modify the Date attribute.

The following table details the attribute settings.

Sample Manager - sample editor parameters with drop-down lists:

Attributes	Description
Sex	This can be set to UNKNOWN, MALE or FEMALE
Condition	This can be set to UNKNOWN, NORMAL, CHALLENGED, PERTURBED, MODIFIED and AFFECTED.
Tag	This is the isotope label used in an Expression Analysis experiment. For samples that are not involved in quantification studies, this value will not be set. While this value can be set using this tool, it is more appropriate to set it in the Expression Analysis tool.
Databank Hyperlinks	To attach a databank hyperlink to a sample: 1. Click the Databank field, and then click a database in the list. 2. In the Unique Identifier field, enter the unique identifier of the required databank entry. 3. Click the Save button to add the hyperlink. **Alternative:** Click the New button to save the current hyperlink and create a new row in which another hyperlink can be entered. **Requirement:** For a databank to appear in the list, its URL must be entered as a bookmark (see Bookmarks tab on page 2-11) and set as non-static. Using SWISS-PROT TrEMBL as an example, it is necessary to enter an accession number in the Unique Identifier field to generate a valid hyperlink.

Generating processed samples

Any number of samples can be mixed together to produce a processed sample. Selected samples are automatically generated into processed samples.

Processed samples can be used in Expression Analysis.

To generate a processed sample:

1. Select two or more original samples (use Shift or Ctrl while selecting), and then right-click.
2. Click Generate Processed Sample.

A new sample is produced and added below the Processed Samples node. The samples from which the new processed sample is generated are also listed in the navigator tree. You can annotate the new sample.

5 Specifying samples, vials, and plates with Container Manager

Container Manager is fundamental to ProteinLynx Global SERVER. It enables you to perform a number of operations:

- Specify the samples and data you want to analyze.
- Attach templates that determine how data is processed.
- Start processing.
- Access your results.

Understanding Container Manager is the quickest way to get up and running with PLGS.

Requirement: Specify your instrument before beginning to use Container Manager (see Instrument tab on page 2-10).

Contents:

Topic	Page
What is Container Manager?	5-2
Importing and viewing PLGS sample lists	5-3
Creating a new vial, microtitre or target plate	5-9
Setting a sample	5-11
Attaching raw data	5-13
Processing raw data	5-17
Re-searching processed data	5-20
Adding processing parameters templates	5-21
Exporting and importing mass spectra	5-22
Working with plates	5-23
Simplifying peaks with SuperTrack	5-26
Interfacing with MassLynx	5-29
Troubleshooting failed client-server workflows	5-33

What is Container Manager?

Container Manager can be used to:

- Import lists of samples that you want to process using PLGS, and associate raw data with the samples in those lists.
- Assign raw data to samples that are attached to vials or plates – the data can be processed, searched, and viewed using the PLGS results browser (Chapter 6 - Viewing results in the Results Browser).
- Export sample lists to MassLynx (see Exporting a sample list to MassLynx on page 5-29) – the data is acquired in MassLynx (see Acquiring data on page 5-31) and the results viewed in the PLGS results browser.

See also: For an explanation of what the term 'sample' means within PLGS, and how samples are used, see Chapter 4 - Annotating and tracking samples with Sample Manager.

To open Container Manager, click the Container Manager icon in the tool tray.

Workflow templates and Processing parameters

The following sections refer to workflow templates and processing parameters:

- Workflow templates – used to perform an automated databank search of samples.
- Processing parameters – determine how the raw spectrum data are processed and whether certain attributes (for example, smoothing) are considered.

For more information on these concepts, including information on how to create your own workflow templates and processing parameters, see Defining templates for searching with Workflow Designer on page 7-1 and Creating custom processing parameters on page 8-1.

Importing and viewing PLGS sample lists

Sample lists can be used to organize the samples you want to work with. You can create a list of samples to be processed using ProteinLynx Global SERVER, and then import that list into PLGS.

Rule: PLGS sample lists – tab- or comma-delimited text files – are different from MassLynx sample lists.

Sample lists are one way of organizing the samples you want to work with: you might find them more convenient than identifying samples by vial, microtitre plate, or target plate.

Importing PLGS sample lists

Requirements: Certain requirements apply to sample lists that you intend to import. For details see Sample list requirements on page 5-4.

To import a sample list:

1. In the navigator tree, click Sample Lists, and then right-click.
2. Click Import Sample List.
3. In the Sample List Chooser dialog box, browse to the sample list file you wish to import, and then click Open.
4. Type a title for the sample list. This title is the name that is displayed within ProteinLynx Global SERVER.

Results:

- The imported sample list is added to the navigator tree, under Sample Lists.
- The samples specified in the list are added under a node that bears the title you specified when you imported the list.
- The contents of the list are displayed in the right-hand side of Sample Manager.
- The samples are added to the Sample Manager tree (see Annotating and tracking samples with Sample Manager on page 4-1).

Sample list requirements

Rule: MassLynx sample lists are not suitable for importing into PLGS.

There are requirements for any sample list that you will import into PLGS:

- It must be a text file.
- Columns must be either comma-separated or tab-separated.
- If columns are comma-separated, the file extension must be .csv. If columns are tab-separated, the file extension must be .txt.

Two columns must appear in the sample list: Sample Name and Data Path.

Required columns in sample lists:

Column name	Description
Sample Name	The name of the sample. It can be either an existing sample in the current project or a completely new sample.
Data Path	The path to either a raw data folder or a processed data file (.xml, .pkl, or .txt).

Additionally, PLGS recognizes several other columns, which you can optionally include in the sample list.

Optional recognized sample list columns:

Column name	Description
Raw Data Location	If the Data Path column refers to raw data paths then this column will be the IP address or name of the computer the raw data is located on. If this column is not present in the sample list then it is assumed the raw data is located on the local machine.
Workflow Template	The name of an existing workflow template in the current project, or the path to an XML workflow template file.
Processing Parameters Template	The name of an existing processing parameters template in the current project, or the path to an XML processing parameters template file.

Optional recognized sample list columns:

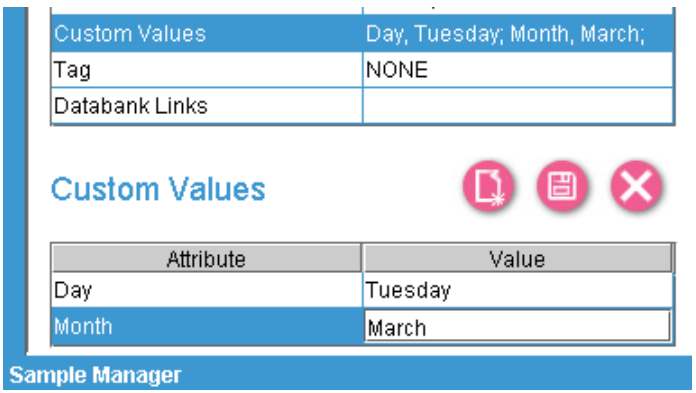
Column name	Description
Parent Sample	The presence of two or more Parent Sample columns indicates that the sample referred to in the Sample Name column is a processed sample. This column can contain the name of a sample in the current project, or a new sample.

Any sample attribute that appears, and is modifiable, in Sample Manager (see Annotating and tracking samples with Sample Manager on page 4-1) can be specified through the inclusion of a column in the sample list.

Example: If an imported sample list includes a column named Time Point, the Time Point attribute of any sample specified in that sample list is set to the value in the sample list column.

Any column header that does not match a sample attribute, or one of the column headers in the tables above, is interpreted as a custom value. Custom values are associated with the sample, and can be viewed and modified using Sample Manager.

Example custom values in Sample Manager:



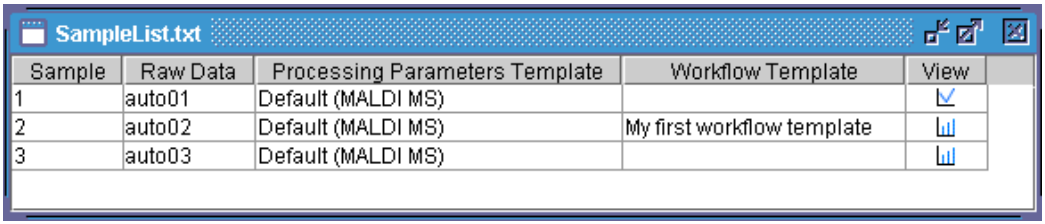
Viewing PLGS sample lists

Once a sample list has been imported, you can view the list and modify certain aspects of it. You can also use the list to view the spectra and workflow results associated with a sample.

The sample list table provides an alternative to the navigator tree for viewing, editing, and processing the data in a sample list.

To open the table for a sample list, click the sample list in the navigator tree, right-click, and then click View Sample List Table.

Sample List table:



Data, either raw or processed, that is associated with a sample in the sample list is represented as a single row in the table.

There are several columns in a sample list table.

Sample list table columns:

Column name	Description
Sample	The name of the sample.
Raw Data	The name of the raw data. Cells in this column have tool tips that display the full path to the raw data, where appropriate.
Processing Parameters Template	The name of the processing parameters template attached to the raw data. If the data represented by a row is processed, this column is empty.
Workflow Template	The name of the workflow template most recently attached to the data. If there is no workflow template attached to the data, this column is empty.
View	An icon that indicates the status of the data. The icon also provides access to the processed spectrum view and the latest workflow results.

View column

The view column contains an icon indicating the status of the data represented by the row. Depending on the status, clicking the icon displays the processed spectrum or workflow results.

View column icons:

	Indicates that the data represented by the row has not been processed.
	Indicates that the data represented by the row is processed data, or raw data that is newly processed. Clicking this icon displays the processed spectrum. **Rule:** If the row represents raw data that has been processed several times, the processed spectrum displayed is for the most recently-processed data.
	Indicates that the data represented by the row has workflow results available. Clicking this icon displays the workflow results for the most recently-submitted workflow. **Rule:** If the row represents raw data that has been processed several times, the most recent workflow results for the most recently-processed data are displayed.

Processing and Searching

To process and search data from the sample table:

1. Click the row representing the data you wish to process. To select multiple rows, hold Shift or Ctrl while clicking.
2. Right-click, and then click on one of these options:
   - Click Process Raw Data to submit the selected raw data for processing and then run the most recently-attached workflow template.
   - Click Process Mass Spectrum to run the most recently-attached workflow template for the selected processed data.

Changing Templates

The processing parameters template associated with data can be changed in the sample list table, and workflow templates added.

To change processing parameters or add workflow templates:

1. Click the row representing the data you wish to change or add a template to. To select multiple rows, hold Shift or Ctrl while clicking.
2. Double-click a cell in the Processing Parameters Template or Workflow Template column, depending on which template setting you want to modify.
3. Click the template you wish to associate with the selected data from the drop-down list.

   Tip: If the template you want to use is not displayed in the list, click the last item – Choose new Processing Parameters / Workflow Template from file – then browse to the desired template.

Result: All the selected rows are updated with the new selection.

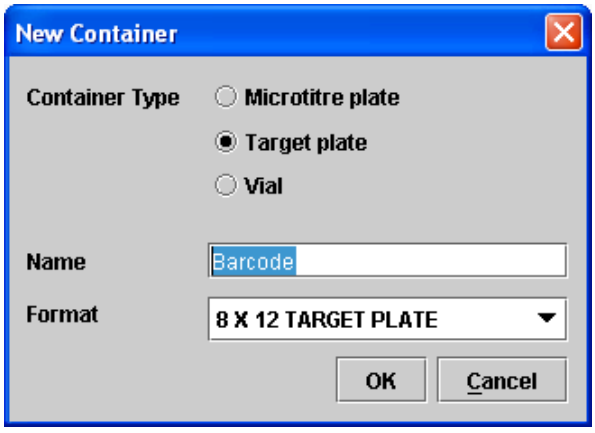
Creating a new vial, microtitre or target plate

The following section describes the creation of a target plate. The process for creating a new vial or microtitre plate is similar.

To create a new target plate:

1. In the navigator tree, click Target Plates, and then right-click.
2. Click New Target Plate.

   New Container dialog box:

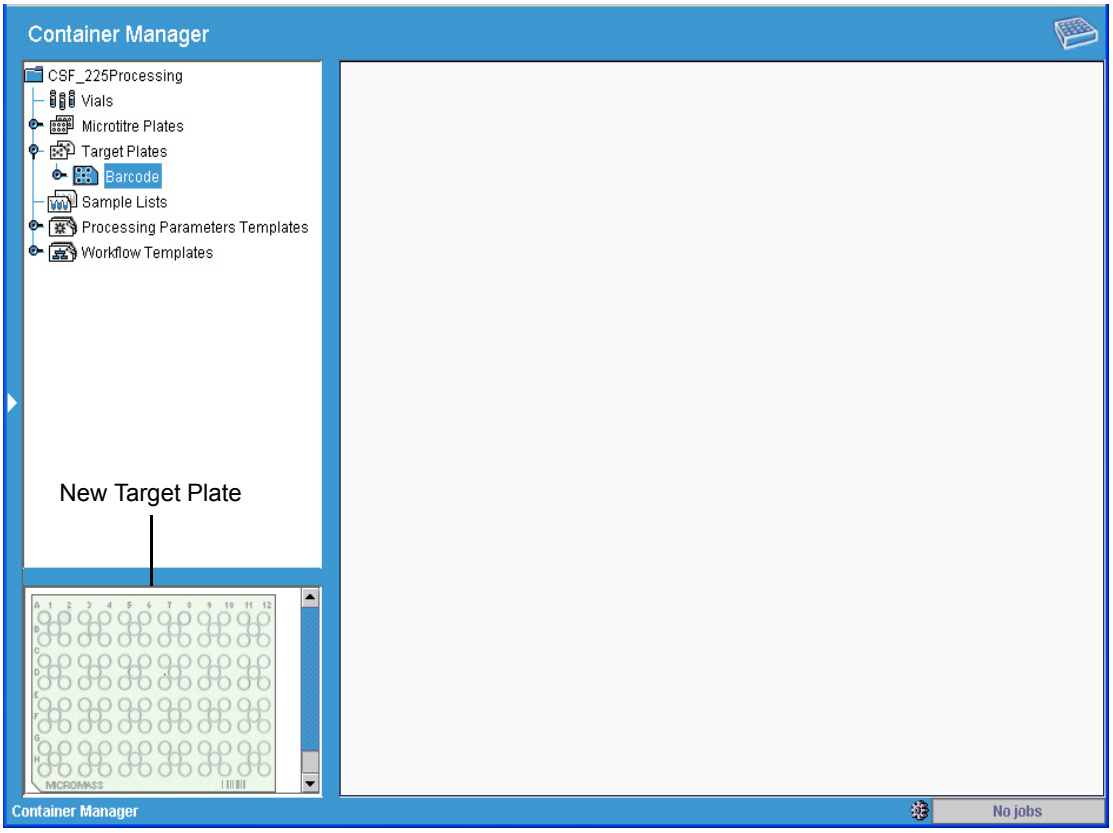


3. In the Barcode text box, type a title or identifying number.
4. If required, select a format for the plate.
5. Click OK.
6. In the navigator tree, expand the Target Plates node, and then click the new plate.

   Result: Two new displays open:

   - The Plate Viewer below the navigator tree displays a graphic of a target plate.

New target plate display:



Setting a sample

See also: For details about how to create samples, see Annotating and tracking samples with Sample Manager on page 4-1.

If a vial, microtitre plate, or target plate is being used, the vial or plate must be associated with a PLGS sample manually. If a sample list was imported, each data file – whether raw or processed – is already associated with a sample.

To set the sample:

1. Open the Select a Sample dialog box, following the instructions in the following table.

   Setting samples:

For this type of container	Do this
Vial	1. Click the vial you wish to set the sample for. 2. Right-click, and then click Set Sample.
Microtitre plate	1. Click the microtitre plate you wish to set samples for. 2. Click a spot on the microtitre plate display. 3. Right-click, and then click Set Sample.
Target plate	1 Click the target plate you wish to set samples for. 2. Click a spot on the target plate display. 3. Right-click, and then click Set Sample.

2. In the Select A Sample dialog box, click Default, and then click OK.

   Tip: Sample Manager (see Annotating and tracking samples with Sample Manager on page 4-1) enables you to organize and annotate your samples. If you have already created samples in Sample Manager, you

will be able to choose them at this stage, and then track and use them throughout your PLGS project.

Result: A new node is added to the navigation tree, below the container selected. If a sample has been set for a microtitre or target plate spot, the spot changes color.

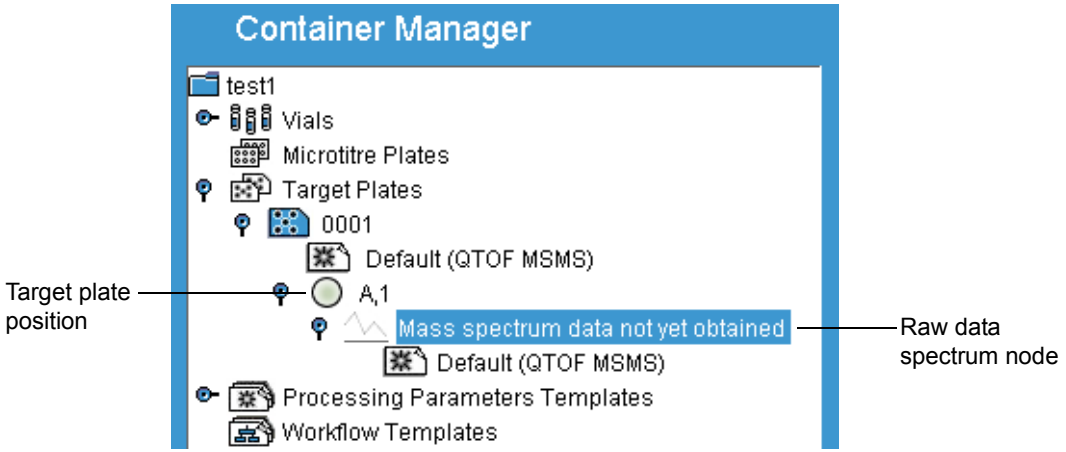
Attaching raw data

If a vial, microtitre plate, or target plate is being used, the raw data must be attached manually. If a sample list was imported, the raw or processed data is already attached to those samples.

To select raw data:

1. In the Container Manager navigator tree, click the Raw Data Spectrum Node, and then right-click.

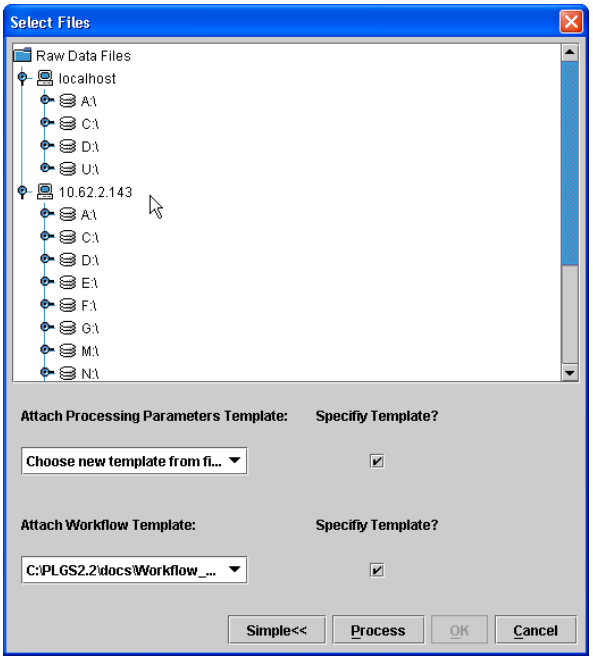
Navigator tree: Mass spectrum data not yet obtained:



In this example, the instrument QTOF MSMS has been set already. See Instrument tab on page 2-10 for information on how to change this.

2. Click Set Raw Data File.

Select Files dialog box for single well - Advanced:



3. Select a raw data file from either the local machine or a remote processor.

   Rule: You can only select one file.

4. Click Advanced to display additional options where you can specify the workflow and processing parameters templates, and also process the data.

5. If you do not intend to process the data immediately, click OK.

   Result: The file name is displayed in the Raw Data Spectrum Node.

Selecting more than one well or spot

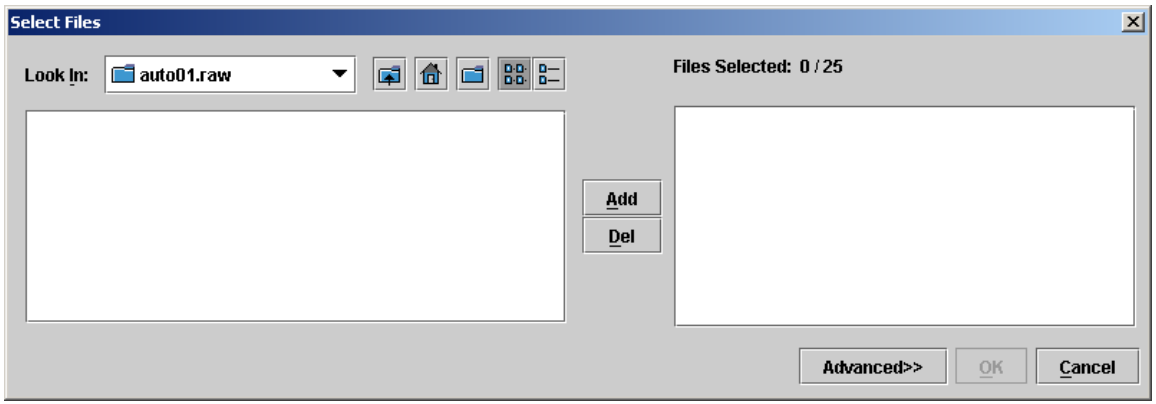
When setting the raw data, it is possible to select data for multiple wells or spots. However, only one raw data file can be attached to each well or spot.

To select more than one well:

1. Click and drag around the wells in the Target Plate (see Figure titled “New target plate display:” on page 5-10) to import data.

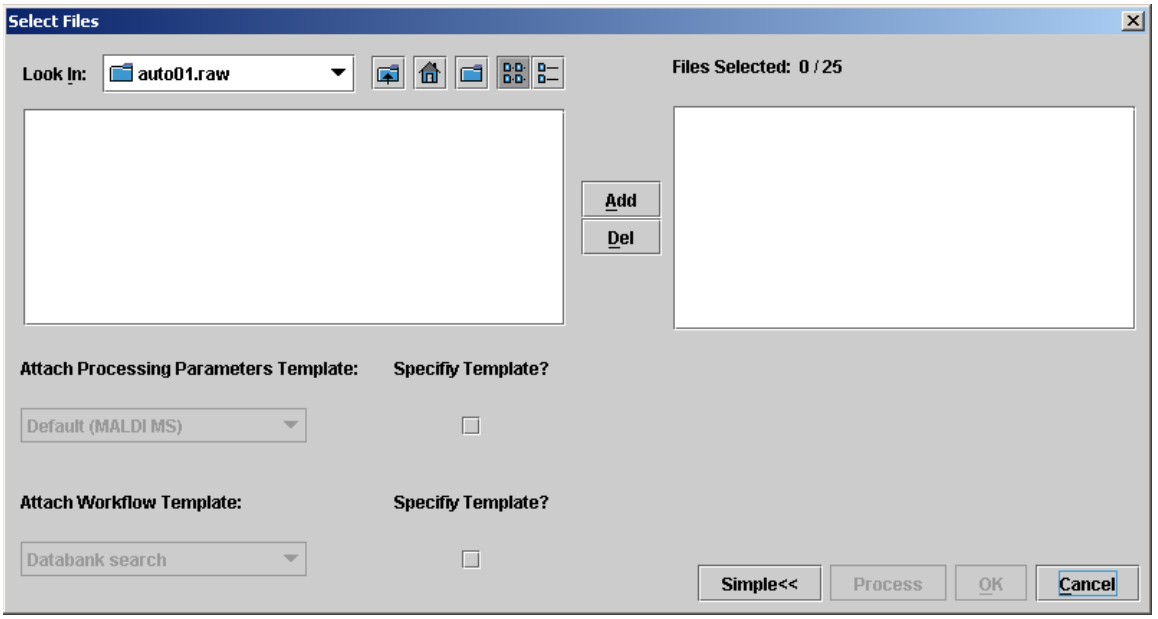
2. Right-click, and then click Set Raw Data File.

Select Files dialog box for multiple files - simple:



3. Select the required raw data files in the left-hand pane from either the local machine or a remote processor, and then click Add. To select multiple files, hold Shift or Ctrl while clicking.
4. Click Advanced to display additional options, in which you can specify the workflow and processing parameters templates, and also process the data.

Select Files dialog box for multiple files - advanced:



The dialog box regulates the number of files attached to wells or spots.

Example: If you select nine files and there are six wells, only the first six files selected are attached to the wells. If you select six files and there are nine wells, files are attached only to the first six wells.

If a well or spot already contains raw spectrum data, a dialog box opens to give you the option to replace the existing raw data. However, if the raw data has been sent for processing it cannot be replaced; a warning message is displayed.

Processing raw data

1. To process the data from the navigator tree, click the Raw Data Spectrum Node, and then right-click.
2. Click Attach Workflow Template, and then click OK to choose a new workflow template from file.

   Tip: You might not need to do this if a workflow template was specified in an imported sample list.
3. Browse to a workflow template, and then click Open. The template is displayed in the navigator tree.

   Rule: Do not attach a PMF workflow template to Electrospray High/Low data.

   See also: For more information on workflow templates and how to produce them, see Chapter 7 - Defining templates for searching with Workflow Designer.
4. Click the Raw Data Spectrum Node again and right-click.
5. Click Process.

As the data is processed, the icons change for the workflow and spectrum (see Workflow and spectrum icons in the navigator tree on page 5-18). Also, the color of each sample well updates according to the search results (see Customizing the plate view on page 5-25).

To view the results, do one of the following actions:

- In the navigator tree, click the name of the workflow.
- In the Results Summary table, click the relevant row.

For details about the results display, see Chapter 6 - Viewing results in the Results Browser.

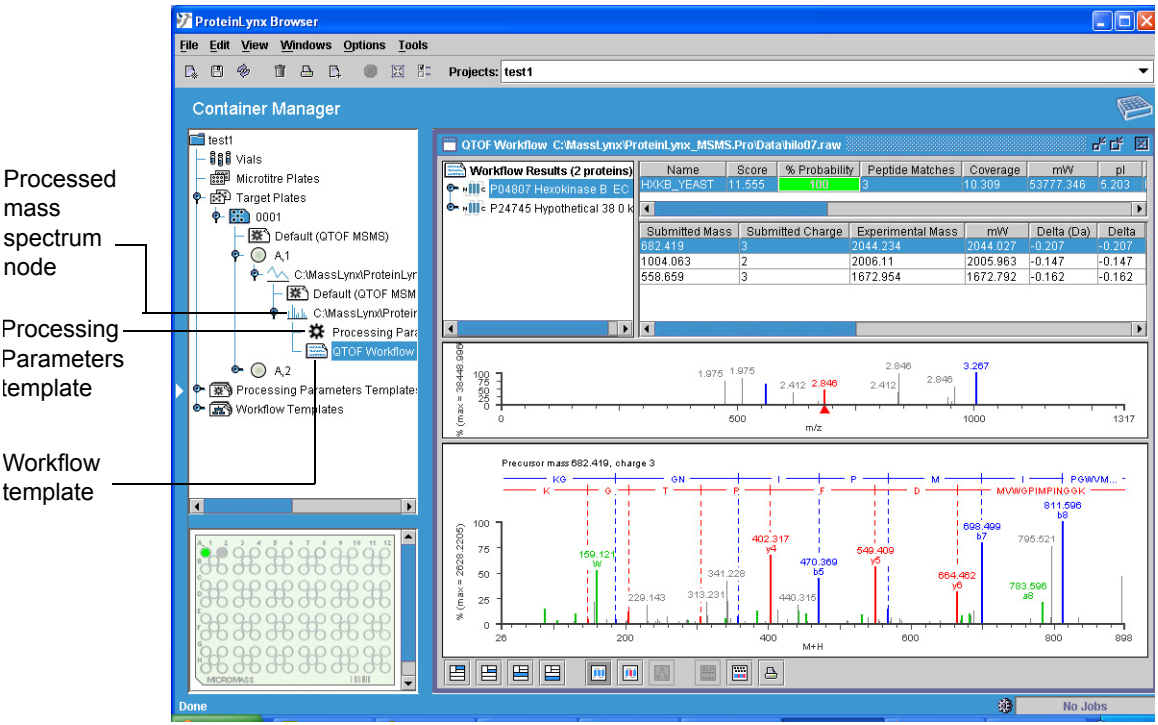
Workflow and spectrum icons in the navigator tree

As the raw data is processed, the icons displayed in the navigator tree change to indicate the progress of the workflow.

Navigator tree processing icons:

Icon	Description
	No raw data is attached to the mass spectrum node.
	Unprocessed data is attached to the mass spectrum node.
	Processed data is attached to the mass spectrum node. **Rule:** Applies to data processed in the browser or imported as an XML file.
	Processed data that has been successfully lockmass corrected is attached to the mass spectrum node.
	Data that has been processed with SuperTrack is attached to the mass spectrum node.
	A workflow template is attached but not processed.
	Processing of the workflow template has failed. See Troubleshooting failed client-server workflows on page 5-33.
	Processing of the workflow template is in progress.
	Processing of the workflow template is complete, but has partially failed.
	Processing of the workflow template is complete. Click to view results (Browser displaying processed data: on page 5-19).

Browser displaying processed data:



Viewing the mass spectrum

Data from a processed mass spectrum node can be viewed in the Processed Data Viewer.

To view the processed spectrum:

1. Click a processed Mass Spectrum node, and then right-click.
2. Click View Spectrum.

Result: The Processed Data Viewer displays the processed spectrum with a list of corresponding monoisotopic masses.

Re-searching processed data

To add more workflow templates to the processed mass spectrum node:

1. In the navigator tree, click the processed mass spectrum node that you wish to add a workflow template to, and then right-click.
2. Click Attach Workflow Template.
3. Click a workflow template in the drop-down list, or click Choose new workflow template from file.
4. If you have selected to choose a new template, browse to the template in the Select Workflow Template XML File dialog box, and then click Open.
5. Click the new workflow template that has been added to the navigator tree, and then right-click.
6. Click Start Workflow to start the process. A prompt for a workflow title is displayed.
7. Click OK to start the process.
8. To display the results, click the new workflow template.

Adding processing parameters templates

So far, all the processing has been done using the default processing parameters. However, different Processing Parameter Template files can be attached to the Raw Data Spectrum Node of the navigator tree. Once added, all the templates that are part of the project are displayed under the Processing Parameters Templates node.

See also: Processing Parameter Template files are produced with the Data Preparation tool: see Creating custom processing parameters on page 8-1 for details.

To add processing parameter template files:

1. In an unprocessed Raw Data Spectrum Node for a well, click the Processing Parameters Template, and then right-click.
2. Click Change Processing Parameters.
3. In the drop-down list, click either 'Choose new processing parameters template from file', or one of the Processing Templates.

   Rule: The Processing Parameters Templates that appear in the drop-down list are those that are already part of the project and are listed under the Processing Parameters Templates node in the navigator tree.

The new Processing Parameters Template is:

- Changed in the Raw Data Spectrum Node.
- Added to the Processing Parameters Templates node at the bottom of the navigator tree.

Exporting and importing mass spectra

PLGS exports and imports mass spectra in XML file format.

Exporting mass spectra

Any processed spectrum can be exported.

To export a processed spectrum:

1. Click a processed Mass Spectrum node, and then right-click and click Export Spectrum.
2. Type an appropriate file name.
3. Click Save.

Importing mass spectra

Mass spectra saved as an XML file can be imported into PLGS.

To import a mass spectrum:

1. Click 'Mass spectrum data not yet obtained' in the navigator tree (Figure titled "Navigator tree: Mass spectrum data not yet obtained:" on page 5-13), and then right-click.
2. Click Import Mass Spectrum.
3. Browse to an appropriate XML file, and then click Open.

Result: The icon on the Mass Spectrum node changes, indicating that processed data is now attached to it (see the table Figure titled "Navigator tree processing icons:" on page 5-18).

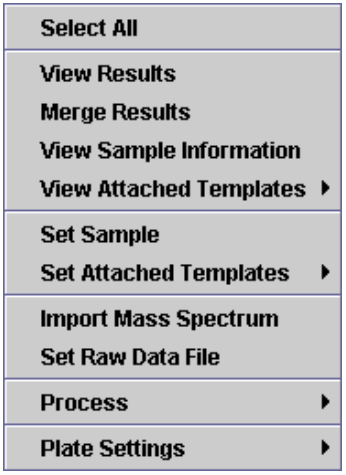
Working with plates

There are several options available in pop-up menus for target plates and microtitre plates. Many of these are the same as the options available from the Container Manager navigator tree.

The available options are the same for target plates and microtitre plates.

To display the Plate menu click a well (or drag across a number of wells), and then right-click.

Plate menu:



You can use the following menu options.

Plate pop-up menu options:

Option	Description
Select All	Selects all the wells on the plate.
View Results	Opens the results browser, see Viewing results on page 6-2.
Merge Results	See Merging MSMS spectra and results on page 5-24.
View Sample Information	Displays sample information on the right-hand panel.
View Attached Templates	Select to display either a workflow template or processing template.

Plate pop-up menu options: (Continued)

Option	Description
Set Sample	Described in Setting a sample on page 5-11.
Set Attached Templates	Set the processing and workflow templates. Each option will open a dialog box in which previously saved templates can be selected.
Import Mass Spectrum	This option is the same as described in Importing mass spectra on page 5-22.
Set Raw Data File	This option is the same as described in Attaching raw data on page 5-13.
Process	Process raw data or latest data.
Plate Settings	See Customizing the plate view on page 5-25.

Merging MSMS spectra and results

If a sample has been separated into several fractions prior to being mass analyzed (such as in a 2D LC or MudPIT experiment), it can be preferable to merge the results that are generated from these fractions.

See also: For further details on samples, see Annotating and tracking samples with Sample Manager on page 4-1.

To merge MSMS spectra and results:

1. Select the required wells or spots, and then right-click.
2. Click Merge Results.
3. Select the sample for which the results need to be merged.
   - Only those samples that are associated with two or more of the selected positions are listed; the default sample is never included.
   - These positions must also contain workflow results generated from Q-Tof-MSMS data.

     Rule: For positions with more than one set of completed workflow results, the most recent will be included in the merge.

Results:

- If the sample selected is associated with a vial, the merged workflow results and data will appear beneath the appropriate vial icon. If the

selected sample has no associated vial, a new one will be automatically added to the current project to act as a place holder for the merged spectra and results.

- The title for the merged results and data is automatically generated and contains the time and date of the merge action.
- The results themselves will be displayed in a workflow results window and have the same format as a single set of workflow results.
- The merged workflow results will not contain duplicate proteins, but all the submitted masses will be included even if they are duplicated.

Customizing the plate view

To modify the colors of the plate view:

1. Click Options > Preferences > Colours tab.

For further details, see Colours tab on page 2-12.

Simplifying peaks with SuperTrack

Rule: SuperTrack is only available for MS^E data.

The SuperTrack tool enables you to validate your raw data before performing databank searches. It looks for replicate EMRTs (Exact Mass Retention Times), and reports only those peaks that have the same m/z and retention time for all three replicates. Further, the high energy peaks must associate with the same precursor in all three cases.

The simplified spectra can accelerate databank searching and improve protein identification. This can be particularly beneficial if you intend to perform databank searching using Mascot, as Mascot prefers fewer peaks. See www.matrixscience.com for more details about Mascot.

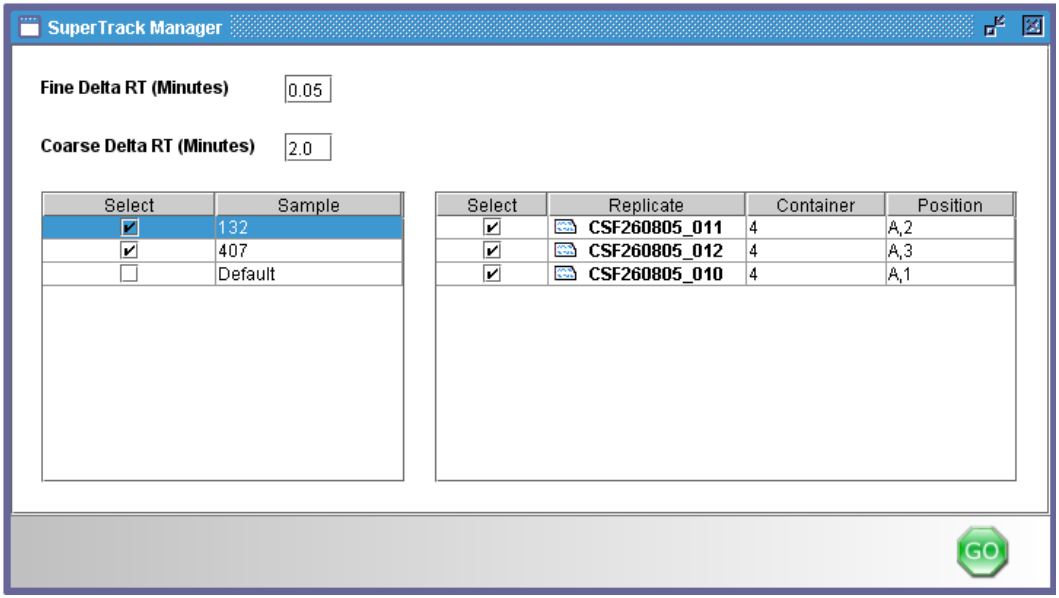
Requirement: Processed data must include retention time information to be compatible with SuperTrack. Data processed with PLGS versions prior to 2.2.5 does not include retention time information.

To open SuperTrack:

1. In the tool tray, click Container Manager.
2. Open a ProteinLynx Global SERVER project by clicking the Projects drop-down box in the toolbar, and then clicking the name of the project.
3. Click Edit > Run SuperTrack.

Result: The SuperTrack Manager is displayed.

SuperTrack Manager:



The SuperTrack Manager provides access to several settings:

- Fine Delta retention time – the retention time tolerance for a replicate, reflecting the precision with which retention time can be estimated within a single function, such as high energy.
- Coarse Delta retention time – the retention time tolerance between replicates, reflecting the reproducibility of retention time across different injections of the same sample.
- Project samples (as defined in Sample Manager – see Annotating and tracking samples with Sample Manager on page 4-1).
- Replicates associated with the selected samples

To run SuperTrack:

1. Select check boxes beside the project samples of interest.
2. Select the check boxes beside the replicates you want to SuperTrack.
3. Click Go.

Result: SuperTrack spectrum nodes appear in the Container Manager tree for each selected sample. Processing can take some time – progress is shown at the bottom right corner of the ProteinLynx browser.

Tip: The same Supertrack spectrum applies to all three replicates of a sample: it is not necessary to perform a databank search on the Supertrack spectrum for each replicate.

To view SuperTrack parameters:

1. Click a SuperTrack spectrum node (see Workflow and spectrum icons in the navigator tree on page 5-18) in the Container Manager tree, and then right-click.
2. Click View SuperTrack Parameters.

Result: The parameters used for SuperTrack processing are displayed. The replicate currently selected in the tree is shown in red.

To view Supertrack spectra:

1. Click a SuperTrack spectrum node in the Container Manager tree, and then right-click.
2. Click View Spectrum.

Exporting SuperTrack results as XML

To export the SuperTrack spectrum as XML:

1. Click a SuperTrack spectrum node in the Container Manager tree, and then right-click.
2. Click Export Spectrum.
3. Browse to a location, and type a name for the XML file to be created.
4. Click Save.

Interfacing with MassLynx

ProteinLynx Global SERVER can export sample lists to MassLynx, where data can be acquired. The data is then imported back into PLGS, where it can be viewed in the results browser.

Exporting a sample list to MassLynx

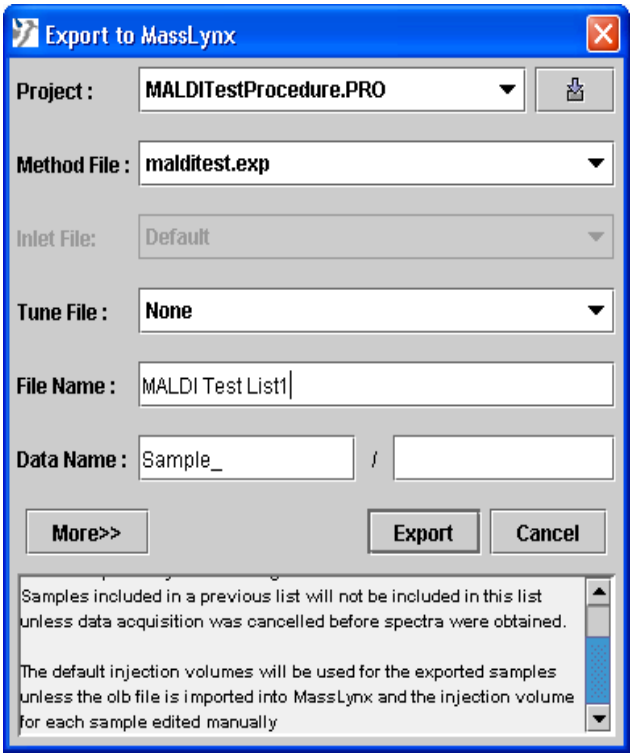
Once samples are set in PLGS (see Setting a sample on page 5-11), but before data is attached to the samples (see Attaching raw data on page 5-13), the samples can be exported to MassLynx as a sample list.

Requirement: Some familiarity with MassLynx is needed. Refer to the MassLynx Online Help for details.

To export a sample list:

1. Right-click the plate or vial node, and then click Export Sample List to MassLynx.

Export to MassLynx dialog box:


Export to MassLynx
Project :
MALDITestProcedure.PRO
Method File :
malditest.exp
Inlet File:
Default
Tune File :
None
File Name :
MALDI Test List1
Data Name :
Sample_
/
More>>
Export
Cancel
Samples included in a previous list will not be included in this list unless data acquisition was cancelled before spectra were obtained.
The default injection volumes will be used for the exported samples unless the olb file is imported into MassLynx and the injection volume for each sample edited manually

2. Select:
   - A Project to export to.
   - An MS Method file from the drop-down list.
   - An appropriate Inlet file (for Q-Tof MSMS only).
   - A Suitable Tune file.
   - A File Name for the MassLynx sample list.
   - An MS Data Name.
3. Click Export.
4. Open MassLynx.
5. Click File > Open Project to open the relevant project.
6. Click File > Import WorkSheet to import the .olb file. Navigate to the relevant MassLynx project and click the .olb file with the name you specified.

7. Click Open.

Result: The MassLynx sample list will be updated.

Acquiring data

Once the sample list is imported into MassLynx, data can be acquired in the normal way.

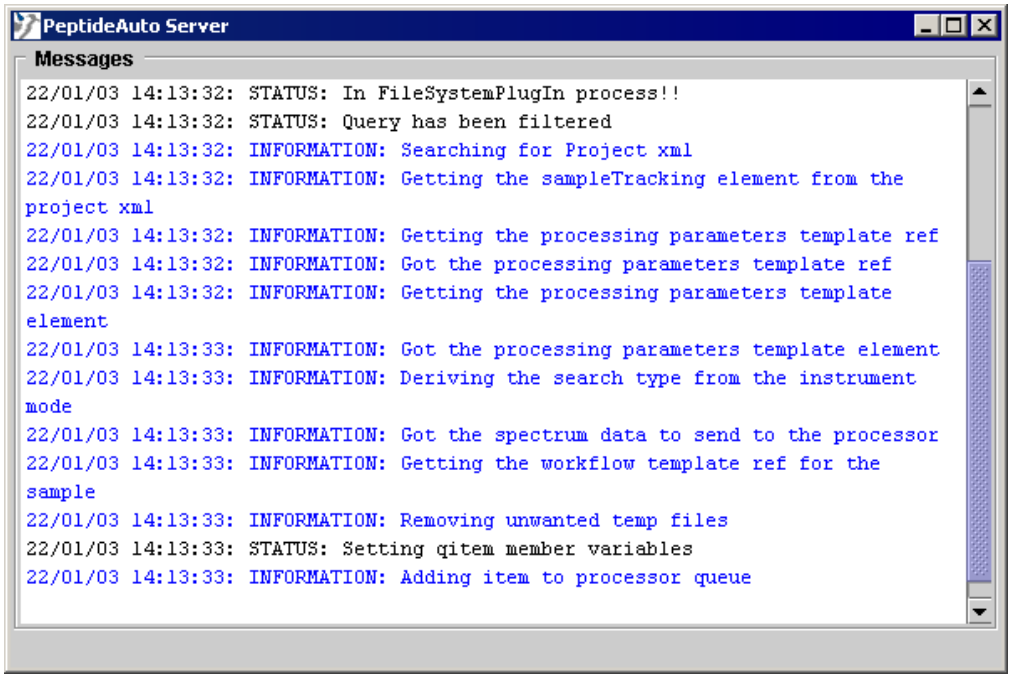
Running the sample list opens the PeptideAuto Server dialog box, which monitors the acquisition.

To acquire data:

1. In the main MassLynx window, click to open the Start Sample List Run dialog box.
2. Select Acquire Sample Data and Auto Process Samples.
3. Click OK.

   The PeptideAuto Server dialog box is opened, which monitors the progress of the acquisition. MassLynx starts to acquire and process data.

PeptideAuto Server dialog box: MassLynx:


PeptideAuto Server
Messages
22/01/03 14:13:32: STATUS: In FileSystemPlugIn process!!
22/01/03 14:13:32: STATUS: Query has been filtered
22/01/03 14:13:32: INFORMATION: Searching for Project xml
22/01/03 14:13:32: INFORMATION: Getting the sampleTracking element from the project xml
22/01/03 14:13:32: INFORMATION: Getting the processing parameters template ref
22/01/03 14:13:32: INFORMATION: Got the processing parameters template ref
22/01/03 14:13:32: INFORMATION: Getting the processing parameters template element
22/01/03 14:13:33: INFORMATION: Got the processing parameters template element
22/01/03 14:13:33: INFORMATION: Deriving the search type from the instrument mode
22/01/03 14:13:33: INFORMATION: Got the spectrum data to send to the processor
22/01/03 14:13:33: INFORMATION: Getting the workflow template ref for the sample
22/01/03 14:13:33: INFORMATION: Removing unwanted temp files
22/01/03 14:13:33: STATUS: Setting qitem member variables
22/01/03 14:13:33: INFORMATION: Adding item to processor queue

4. The data can be viewed periodically in the main PLGS window as it is acquired. To view this data in PLGS, click either:

   - File > Update, or
   - on the toolbar.

   All the latest results are displayed in the browser.

Troubleshooting failed client-server workflows

If workflow queries sent from a client machine are failing (for an example failed workflow icon, see Workflow and spectrum icons in the navigator tree on page 5-18), check the following:

- Check that the client is connected to the correct PLGS server. If you have recently installed the client software, you need to re-add the server using the ProteinLynx Browser Preferences dialog box on the client. For details, see Changing preferences on page 2-5.

  To add a new server to the list, type the IP address in the text field at the top of the dialog box, and then click Apply.

  Any errors displayed are usually because the PLGS server components (search engine/microkernel) are not running on the specified computer.

- Check that the workflows are referencing a databank that exists on the server you are connected to. Check this by opening the workflow template in the Workflow Designer (see Opening workflow templates on page 7-10).

  Check that each databank field contains a databank. If a databank is not shown, the previously set databank is not present on your currently-selected server. This is an issue when opening up older workflows created with a previous version of PLGS.

6 Viewing results in the Results Browser

Following acquisition and processing, the data can be viewed in the workflow results browser. A separate results browser is opened for each set of results.

This section describes how to view results and use the results browser.

Contents:

Topic	Page
Viewing results	6-2
Results browser	6-3
Protein Workpad	6-27
Exclude Masses Workpad	6-31

Viewing results

The results browser for each set of results can be opened in several ways. Each set of results is listed in a Results Summary table.

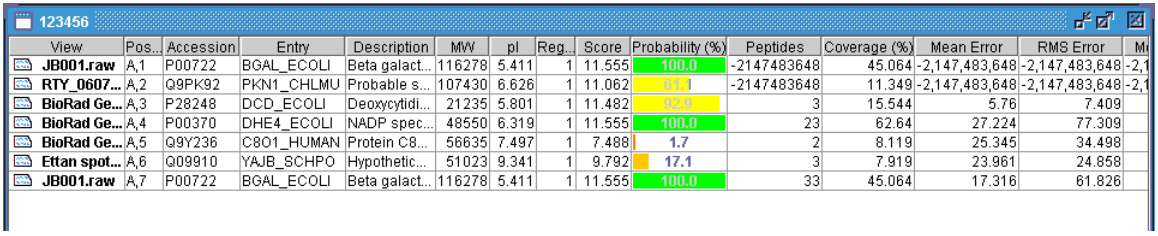
To view results in the results browser, either:

- Click a well or spot on a plate, right-click, and then click View Results, or
- Double-click the workflow results node in the Container Manager navigator tree, or
- Click anywhere in a row of the Results Summary table.

To view a larger Results Summary table:

- Hide the tool tray by clicking the arrow on the blue splitter bar between the tool tray and the display area of the main PLGS window.
- Hide the navigator tree panel by clicking View > Maximise Desktop.

Results Summary table:



123456

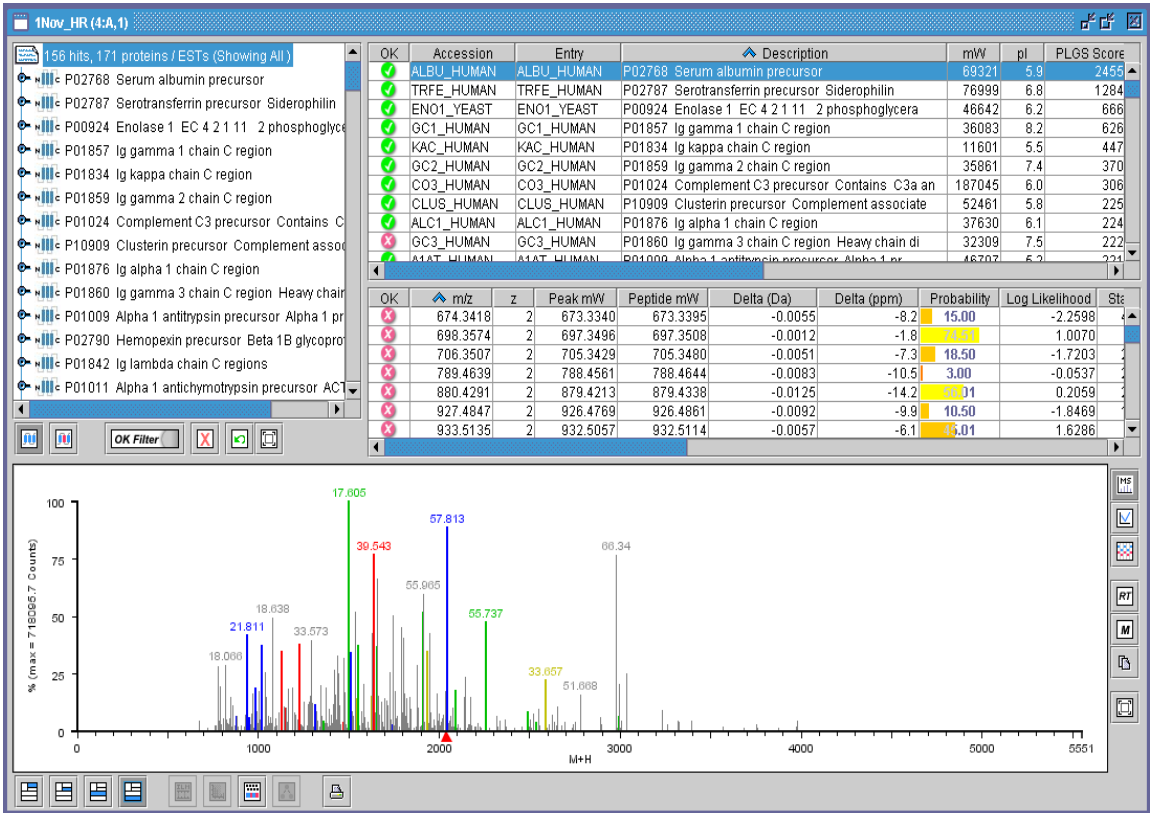
View	Pos...	Accession	Entry	Description	MW	pI	Reg...	Score	Probability (%)	Peptides	Coverage (%)	Mean Error	RMS Error	M
JB001.raw	A,1	P00722	BGAL_ECOLI	Beta galact...	116278	5.411	1	11.555	100.0	-2147483648	45.064	-2,147,483,648	-2,147,483,648	-2,1
RTY_0607...	A,2	Q9PK92	PKN1_CHLMU	Probable s...	107430	6.626	1	11.062	99.1	-2147483648	11.349	-2,147,483,648	-2,147,483,648	-2,1
BioRad Ge...	A,3	P28248	DCD_ECOLI	Deoxycytidi...	21235	5.801	1	11.482	99.9	3	15.544	5.76	7.409	
BioRad Ge...	A,4	P00370	DHE4_ECOLI	NADP spec...	48550	6.319	1	11.555	100.0	23	62.64	27.224	77.309	
BioRad Ge...	A,5	Q9Y236	C8O1_HUMAN	Protein C8...	56635	7.497	1	7.488	1.7	2	8.119	25.345	34.498	
Ettan spot...	A,6	Q09910	YAJB_SCHPO	Hypothetic...	51023	9.341	1	9.792	17.1	3	7.919	23.961	24.858	
JB001.raw	A,7	P00722	BGAL_ECOLI	Beta galact...	116278	5.411	1	11.555	100.0	33	45.064	17.316	61.826	

Adjust the size of any column by clicking and dragging the right-hand side of the column. Change the position of any column by clicking and dragging the column to a new position.

Results browser

The workflow results browser displays mass spectrum data alongside results from Databank searches, AutoMod analyses and De Novo sequencing. The browser can show results from an individual search, or merged results from a workflow containing multiple analyses.

Browser display of results for MS spectrum data:

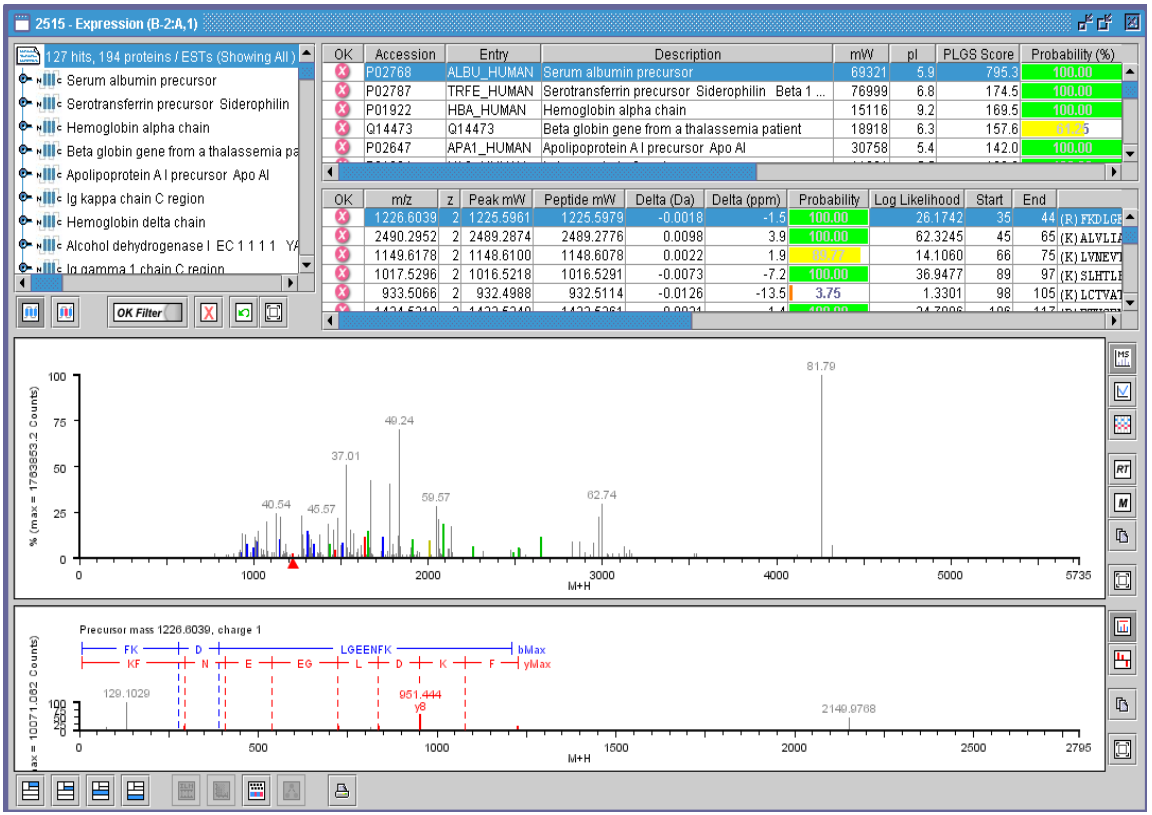

1Nov_HR (4:A,1)
156 hits, 171 proteins / ESTs (Showing All)
P02768 Serum albumin precursor
P02787 Serotransferrin precursor Siderophilin
P00924 Enolase 1 EC 4 2 1 11 2 phosphoglyce
P01857 Ig gamma 1 chain C region
P01834 Ig kappa chain C region
P01859 Ig gamma 2 chain C region
P01024 Complement C3 precursor Contains C
P10909 Clusterin precursor Complement assoc
P01876 Ig alpha 1 chain C region
P01860 Ig gamma 3 chain C region Heavy chain
P01009 Alpha 1 antitrypsin precursor Alpha 1 pr
P02790 Hemopexin precursor Beta 1B glycopro
P01842 Ig lambda chain C regions
P01011 Alpha 1 antichymotrypsin precursor AC
OK Filter
OK | Accession | Entry | Description | mW | pI | PLGS Score
ALBU_HUMAN | ALBU_HUMAN | P02768 Serum albumin precursor | 69321 | 5.9 | 2455
TRFE_HUMAN | TRFE_HUMAN | P02787 Serotransferrin precursor Siderophilin | 76999 | 6.8 | 1284
ENO1_YEAST | ENO1_YEAST | P00924 Enolase 1 EC 4 2 1 11 2 phosphoglycera | 46642 | 6.2 | 666
GC1_HUMAN | GC1_HUMAN | P01857 Ig gamma 1 chain C region | 36083 | 8.2 | 626
KAC_HUMAN | KAC_HUMAN | P01834 Ig kappa chain C region | 11601 | 5.5 | 447
GC2_HUMAN | GC2_HUMAN | P01859 Ig gamma 2 chain C region | 35861 | 7.4 | 370
CO3_HUMAN | CO3_HUMAN | P01024 Complement C3 precursor Contains C3a an | 187045 | 6.0 | 306
CLUS_HUMAN | CLUS_HUMAN | P10909 Clusterin precursor Complement associate | 52461 | 5.8 | 225
ALC1_HUMAN | ALC1_HUMAN | P01876 Ig alpha 1 chain C region | 37630 | 6.1 | 224
GC3_HUMAN | GC3_HUMAN | P01860 Ig gamma 3 chain C region Heavy chain di | 32309 | 7.5 | 222
OK | m/z | z | Peak mW | Peptide mW | Delta (Da) | Delta (ppm) | Probability | Log Likelihood
674.3418 | 2 | 673.3340 | 673.3395 | -0.0055 | -8.2 | 15.00 | -2.2598
698.3574 | 2 | 697.3496 | 697.3508 | -0.0012 | -1.8 | 1.0070
706.3507 | 2 | 705.3429 | 705.3480 | -0.0051 | -7.3 | 18.50 | -1.7203
789.4639 | 2 | 788.4561 | 788.4644 | -0.0083 | -10.5 | 3.00 | -0.0537
880.4291 | 2 | 879.4213 | 879.4338 | -0.0125 | -14.2 | 0.2059
927.4847 | 2 | 926.4769 | 926.4861 | -0.0092 | -9.9 | 10.50 | -1.8469
933.5135 | 2 | 932.5057 | 932.5114 | -0.0057 | -6.1 | 1.6286
% (max = 718065.7 Counts)
M+H
17.605
57.813
39.543
66.34
55.965
18.638
21.811
55.737
33.573
18.066
33.657
51.668

The results display enables you to select various different views of the data. To view further details, click individual results items.

The results browser is divided into four sections: the navigator tree, table of protein and EST data, table of peptide data, and spectrum viewer. Each section can be resized by clicking and dragging the dividers.

If the results are for MSMS spectrum data, two spectrum viewers are included; one shows the parent spectrum, and the other shows fragmentation data.

Browser display of results for MSMS spectrum data:


2515 - Expression (B-2:A,1)
127 hits, 194 proteins / ESTs (Showing All)
Serum albumin precursor
Serotransferrin precursor Siderophilin
Hemoglobin alpha chain
Beta globin gene from a thalassemia pa
Apolipoprotein A I precursor Apo AI
Ig kappa chain C region
Hemoglobin delta chain
Alcohol dehydrogenase I EC 1 1 1 1 YA
Ig gamma 1 chain C region
OK Filter
OK | Accession | Entry | Description | mW | pI | PLGS Score | Probability (%)
P02768 | ALBU_HUMAN | Serum albumin precursor | 69321 | 5.9 | 795.3 | 100.00
P02787 | TRFE_HUMAN | Serotransferrin precursor Siderophilin Beta 1 ... | 76999 | 6.8 | 174.5 | 100.00
P01922 | HBA_HUMAN | Hemoglobin alpha chain | 15116 | 9.2 | 169.5 | 100.00
Q14473 | Q14473 | Beta globin gene from a thalassemia patient | 18918 | 6.3 | 157.6
P02647 | APA1_HUMAN | Apolipoprotein A I precursor Apo AI | 30758 | 5.4 | 142.0 | 100.00
OK | m/z | z | Peak mW | Peptide mW | Delta (Da) | Delta (ppm) | Probability | Log Likelihood | Start | End
1226.6039 | 2 | 1225.5961 | 1225.5979 | -0.0018 | -1.5 | 100.00 | 26.1742 | 35 | 44 (R) FKDLGE
2490.2952 | 2 | 2489.2874 | 2489.2776 | 0.0098 | 3.9 | 100.00 | 62.3245 | 45 | 65 (K) ALVLIA
1149.6178 | 2 | 1148.6100 | 1148.6078 | 0.0022 | 1.9 | 14.1060 | 66 | 75 (K) LVNEVT
1017.5296 | 2 | 1016.5218 | 1016.5291 | -0.0073 | -7.2 | 100.00 | 36.9477 | 89 | 97 (K) SLHTLF
933.5066 | 2 | 932.4988 | 932.5114 | -0.0126 | -13.5 | 3.75 | 1.3301 | 98 | 105 (K) LCTVAT
% (max = 1703653.2 Counts)
81.79
49.24
37.01
40.54
45.57
59.57
62.74
M+H
Precursor mass 1226.6039, charge 1
FK D LGEENFK bMax
KF N E EG L D K F yMax
129.1029
951.444
y8
2149.9768
M+H

Results tree toolbar

The toolbar below the workflow results tree includes controls for switching between protein and peptide views, and also for filtering results to only show those marked in certain ways.

Results browser - results tree toolbar:

Button	Description
	Switch to protein view.

Results browser - results tree toolbar: (Continued)

Button	Description
	Switch to peptide/masses view.
	Filter the results to show only those marked with the indicated symbol.
	Clear all protein and peptide OK assignments, setting all proteins and peptides to not OK – .
	Reset all protein and peptide OK values to their default assignments.
	Copy an image of the protein or peptide tree to the clipboard.

Bottom toolbar

A toolbar at the bottom of the results browser enables you to quickly open windows and switch between views.

Results browser - bottom toolbar buttons:

Button	Description
	View the Protein Results panel.
	View the Peptide Results panel.
	View the MS Spectrum panel.
	View the MSMS Spectrum panel.
	Show the BLAST (Basic Local Alignment Search Tool) results (see BLAST results on page 14-26 for further details).
	Show a web-page containing the original Mascot results. Available if the search was performed against Mascot.

Results browser - bottom toolbar buttons: (Continued)

Button	Description
	Opens the Protein Workpad (see Protein Workpad on page 6-27).
	Open the PepGrab Parameters dialog box. Available if the databank used is indexed for running PepGrab (see PepGrab on page 6-11 for details).
	Prints the results of the workflow.

Spectrum viewer toolbar

A toolbar to the right of the spectrum viewers enables you to switch between spectrum views, and to copy spectrum data.

Results browser - Spectrum viewer toolbar:

Button	Description
	View the MS spectrum.
	View the raw data.
	View the expected fragment ion masses.
	Show the retention times on the X-axis.
	Show masses on the X-axis.
	Copy spectrum data to the clipboard.
	Copy spectrum image to the clipboard.
	View the MSMS spectrum.

Results browser - Spectrum viewer toolbar:

Button	Description
	View MSMS spectrum ion probabilities.

Results browser navigator tree

The top left component of the results browser is a tree for navigating the workflow results.

The two different views of the data are protein view and peptide/masses view. Individual items from the data (such as a single protein or mass) can be selected within the tree, or dragged and dropped from the tree into another component.

To toggle the navigator tree view, click the Protein View and Peptide View buttons below the tree.

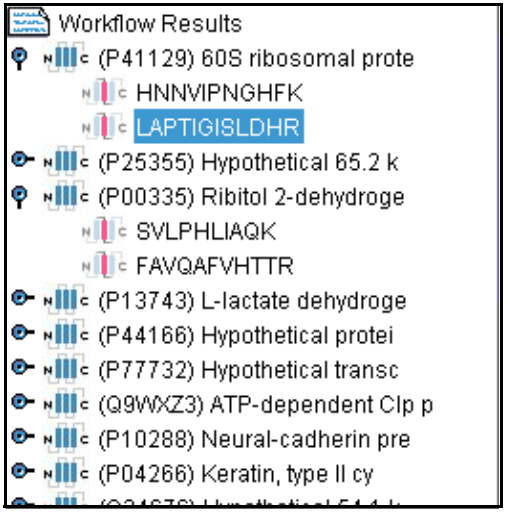
If a workflow contains a BLAST Query then an additional BLAST View is available. The BLAST view – which is accessible by right-clicking the navigator tree, and then clicking Show Blast Results – does not alter the navigator tree; it triggers the display of a BLAST results panel (see BLAST results on page 14-26 for further details).

Protein view

The Protein view displays the proteins and ESTs that were matched to the spectrum data by the analyses. Proteins and ESTs are grouped into hits (each hit represents a set of proteins and ESTs that share the same peptides).

The following illustration shows a typical Protein view.

Navigator panel - Protein view:



The following table details the icons in the Protein and Peptide views.

Navigator panel icons - Protein and Peptide views:

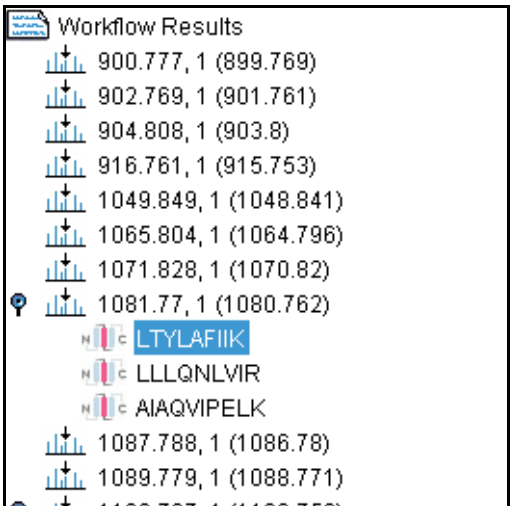
Icon	Description
	Represents a protein or EST. Icons nested directly underneath the Workflow Results icon represent the highest scoring protein or EST for each hit. Further proteins and ESTs can be nested within each hit.
	Represents a peak mass from the mass spectrum.
	Represents a peptide. Peptides are nested underneath the protein or EST to which the peptide sequence has been matched.
	Represents a peptide with post-translational modifications. Peptides are nested underneath the protein or EST to which the peptide sequence has been matched.

Peptide view

The Peptide view displays:

- masses from the spectrum that were used as queries for the search.
- peptides that were matched to the masses.

Navigator panel - Peptide view:



Selecting items in the navigator tree

To select any item in the navigator tree, click the node that represents the item. The other components in the results browser update automatically to reflect the selection.

Selecting one item can cause other items to be selected.

Example: If a peptide is selected, the hit, protein, and peak mass to which the peptide is matched are also selected.

Results of selecting navigator tree nodes:

Icon Selected	Result
Workflow results	All selections are reset. The protein table shows the top-scoring protein or EST from each hit. The peptide table shows the peptides matched to all of the top-scoring proteins and ESTs. The MS spectrum display will color the peaks matched to peptides from the top-scoring protein or EST in the results. The MS/MS spectrum display will show fragmentation data for the first peptide in the peptide table.
Protein or EST	The protein table shows all proteins and ESTs that belong to the same hit as the selection, and the row showing the selected protein or EST is highlighted. The peptide table shows all peptides that have been matched to the selected protein or EST. The MS spectrum display colors the peaks matched to peptides from the selected protein or EST. The MS/MS spectrum display is unchanged.
Peak mass	The protein table is unchanged. The peptide table is unchanged. The MS spectrum display highlights the peak mass. The MS/MS spectrum display shows the fragmentation spectrum for the selected peak mass.
Peptide	The protein table shows all proteins and ESTs that belong to the same hit as the peptide, and the row showing the protein or EST that is matched to the selected peptide is highlighted. The peptide table shows all peptides that have been matched to the same protein or EST as the selection, and the row showing the selected peptide is highlighted. The MS spectrum display highlights the peak mass that is matched to the peptide. The MS/MS spectrum display shows the fragmentation spectrum for the peak mass that is matched to the peptide, and annotates the spectrum with the peptide fragmentation data.

Items can be dragged and dropped onto other components. An example of when this might be useful is when selecting a sequence for a one-off AutoMod query.

PepGrab

You can search a selected databank for peptides that match a given mass, within a set mass tolerance. This enables you to evaluate the quality of a peptide assignment for a given mass and to compare this peptide with others found in the databank for that mass.

Tip: PepGrab is only available if the databank specified in the workflow template that produced the results was set to Index for PepGrab. For details on setting databank attributes, see Databank attributes on page 13-4.

To use PepGrab:

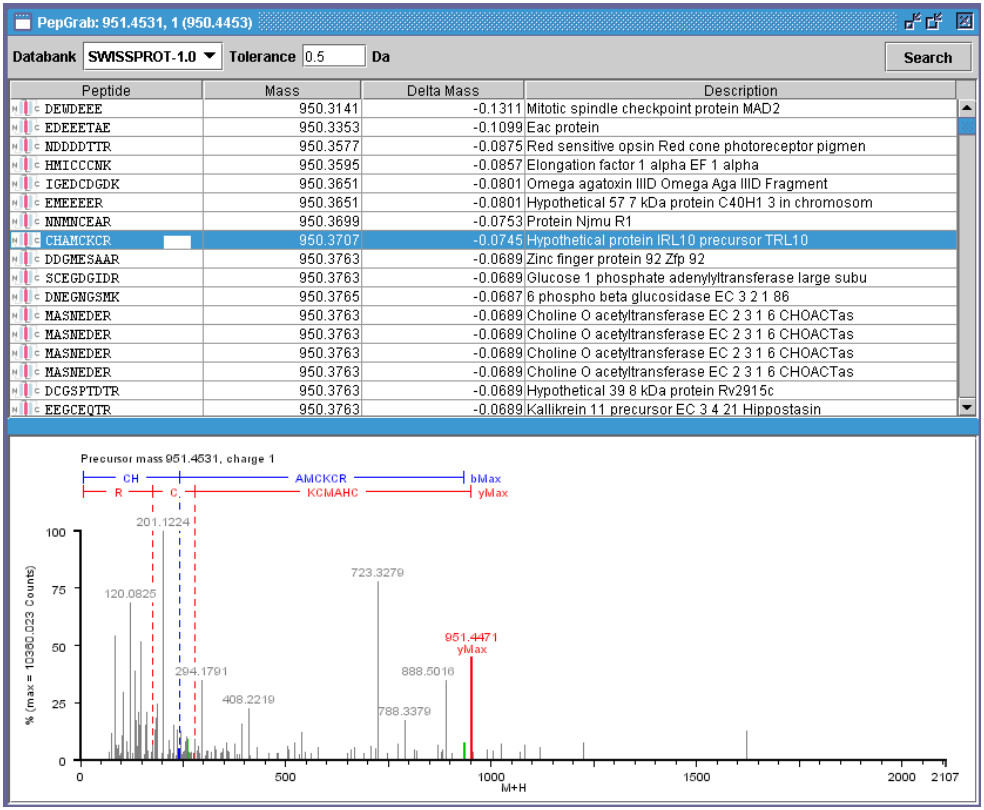
1. In the results table, click a peptide, and then right-click.
2. Click Perform PepGrab.
3. In the list, click a databank to search.
4. Type a mass tolerance (default = 0.5 Da).
5. Click Search.

Result: A list of peptides that match the mass tolerance is displayed.

You can scroll through the list and compare the quality of the fragmentation data for each peptide in the list.

Rule: You cannot replace the original peptide assignment with one of the new assignments returned by PepGrab.

Peptide matches for given mass:


PepGrab: 951.4531, 1 (950.4453)
Databank SWISSPROT-1.0 Tolerance 0.5 Da Search
Peptide Mass Delta Mass Description
DEWDEEE 950.3141 -0.1311 Mitotic spindle checkpoint protein MAD2
EDEEETAE 950.3353 -0.1099 Eac protein
NDDDDTTR 950.3577 -0.0875 Red sensitive opsin Red cone photoreceptor pigmen
HMICCCNK 950.3595 -0.0857 Elongation factor 1 alpha EF 1 alpha
IGEDCDGDK 950.3651 -0.0801 Omega agatoxin IIID Omega Aga IIID Fragment
EMEEEEER 950.3651 -0.0801 Hypothetical 57 7 kDa protein C40H1 3 in chromosom
NNMNCEAR 950.3699 -0.0753 Protein Njmu R1
CHAMCKCR 950.3707 -0.0745 Hypothetical protein IRL10 precursor TRL10
DDGMESAAR 950.3763 -0.0689 Zinc finger protein 92 Zfp 92
SCEGDGIDR 950.3763 -0.0689 Glucose 1 phosphate adenylyltransferase large subu
DNEGNGSMK 950.3765 -0.0687 6 phospho beta glucosidase EC 3 2 1 86
MASNEDER 950.3763 -0.0689 Choline O acetyltransferase EC 2 3 1 6 CHOACTas
MASNEDER 950.3763 -0.0689 Choline O acetyltransferase EC 2 3 1 6 CHOACTas
MASNEDER 950.3763 -0.0689 Choline O acetyltransferase EC 2 3 1 6 CHOACTas
MASNEDER 950.3763 -0.0689 Choline O acetyltransferase EC 2 3 1 6 CHOACTas
DCGSPTDTR 950.3763 -0.0689 Hypothetical 39 8 kDa protein Rv2915c
EEGCEQTR 950.3763 -0.0689 Kallikrein 11 precursor EC 3 4 21 Hippostasin
Precursor mass 951.4531, charge 1
CH AMCKCR bMax
R C KCMAHC yMax
201.1224
120.0825
294.1791
408.2219
723.3279
788.3379
888.5016
951.4471
yMax
% (max = 10360.023 Counts)
M+H

Protein and EST table

The top right component of the results browser is a table that displays a list of proteins and ESTs. Each row in the table represents a single protein or EST, and each column in the table represents a particular data item (for example, accession number).

The first column in the table indicates whether the protein match has been set as good (OK,), possible (Maybe,), or poor (Not OK,). These assignments are either made manually – by clicking in the column to cycle through the options – or automatically during searching. For details on how and when the assignments are made automatically, see Automatic data curation on page B-7.

Tip: Modifying the assignment for a protein or EST will affect the assignments of its associated peptides.

Protein/EST table:

OK	Accession	Entry	Description	mW	pI	PLGS Score	Probability (%)	Peptides	Coverage (%)	Mean Err
✓	P02071	HBB_BOSGF	Hemoglobin beta chain	15976	6.8	5.7	0.28	7	57.9	
✓	P02070	HBB_BOVIN	Hemoglobin beta chain	15944	7.5	5.7	0.27	8	58.6	
?	P18990	HBB_TURTR	Hemoglobin beta chain	16025	6.8	5.2	0.18	3	24.0	
✗	P02067	HBB_PIG	Hemoglobin beta chain	16024	7.6	5.0	0.14	2	15.1	

When the table is initially displayed, or if the Workflow Results icon in the navigator tree is selected, the table shows the highest-scoring protein or EST from each hit in the results. When a hit is selected, the table shows all of the proteins or ESTs that belong to the selected hit.

The following operations can be performed using this table:

- The columns to be displayed, the order of columns, and the precision with which numbers are shown can be controlled.
- Individual proteins and ESTs can be selected in the table, or dragged and dropped into another component.

Peptide table

The middle-right component of the results browser is a table that displays a list of peptides. Each row in the table represents a single peptide, and each column in the table represents a particular data item (molecular weight, for example).

The first column in the table indicates whether the peptide match has been set as good (OK, ✓), possible (Maybe, ?), or poor (Not OK, ✗). These assignments are either made manually – by clicking in the column to cycle through the options – or automatically during searching. For details on how and when the assignments are made automatically, see Automatic data curation on page B-7.

Tip: Modifying the assignment for a peptide will affect the assignments of its associated proteins or ESTs.

Peptide table:

OK	m/z	z	Peak mW	Peptide mW	Delta (Da)	Delta (ppm)	Probability	Start	End	Sequence	
✓	950.5081	1	949.5003	949.5021	-0.0018	-1.9	0.39	8	16	(K)AAVTAFWGK(V)	PLG
✗	1274.7412	1	1273.7334	1273.7183	0.0151	11.9	0.00	30	39	(R)LLVVYPWTQR(F)	PLG
✗	2090.0291	1	2089.0212	2088.9463	0.0750	35.9	0.33	40	58	(R)FFESFGDLSTADAVMNNPK(V)	PLG

When the table is initially displayed, or if the Workflow Results icon in the navigator tree is selected, the table shows all of the peptides from each hit in the results. When a hit is selected, the table shows all of the peptides that belong to the selected hit. When a protein or EST is selected, the table shows all of the peptides that belong to the selected protein or EST.

The following operations can be performed using this table:

- The columns to be displayed, the order of columns, and the precision with which numbers are shown can be controlled.
- Individual peptides can be selected in the table, or dragged and dropped into another component.

Controlling the columns in the tables

To add or remove columns in the tables:

1. Right-click the table.
2. Click Select Table Columns.
3. To add or remove a single column, select or clear the check box for the column on the menu.

   To add or remove multiple columns, click Add/Remove Columns, and then select or clear the check boxes for the relevant columns. Click OK.

To change the order of columns:

Either:

1. Drag and drop the column headers in the table.

Or

1. Right-click the table.
2. Click Select Table Columns.
3. Click Edit Order/Precision.
4. In the Edit Column Order/Precision dialog box, click the column you want to move, and then click the up or down arrow. Repeat for other columns.
5. Click the X in the top right of the dialog box to close.

To change the precision with which numbers are displayed:

1. Right-click the table.
2. Click Select Table Columns.
3. Click Edit Order/Precision.
4. In the Edit Column Order/Precision dialog box, locate the column you wish to modify. The number of decimal places currently displayed for that column is displayed alongside the column name.
5. Click the up or down arrows beside the number. Increasing the number results in more decimal places being displayed; decreasing the number results in fewer decimal places being displayed.
6. Click the X, in the top right of the dialog box, to close.

Selecting proteins and ESTs from the table

To select a protein or EST from the table, click the relevant row.

The peptide table shows all peptides that have been matched to the selected protein or EST. The MS spectrum display highlights the peaks matched to peptides from the selected protein or EST.

Hold down the left mouse button to drag and drop the protein or EST onto another component.

Selecting peptides from the table

To select a peptide from the table, click the relevant row.

The MS spectrum display highlights the peak mass that is matched to the peptide. The MSMS spectrum display shows the fragmentation spectrum for the peak mass that is matched to the peptide, and annotates the spectrum with the peptide fragmentation data.

Hold down the left mouse button to drag and drop the peptide onto another component.

Resubmitting the search

The spectrum data, with some peaks excluded, can be resubmitted for a search. The resubmitted search uses the same query parameters as the search that produced the original set of results.

- To resubmit the unmatched peaks from the spectrum for a search (that is, excluding all peaks already matched to a peptide), right-click either the protein or peptide table, and then click Exclude/Re-submit > Resubmit with Current Exclude List.
- To resubmit all peaks not specifically excluded from the spectrum, right-click either the protein or peptide table, and then click Exclude/Re-submit > Resubmit Excluding Current Protein.

Peaks can be excluded from resubmitted searches using the Exclude Masses Workpad, described in Exclude Masses Workpad on page 6-31.

Note: Masses selected for exclusion are usually theoretical masses, which can differ from masses found in the data. Therefore, due to the possibility of misassignment (a detected mass being mistaken for a different theoretical mass), the corresponding data is suppressed according to how well the masses match the theoretical masses rather than being completely extinguished.

Copying data

To copy the data in a table to the clipboard, right-click either the protein or peptide table, and then click Copy Table Data.

The data copied to the clipboard is organized by row. Each line of copied text represents a single row: the line lists the row number and the data values from the table. Separate data values are comma-separated.

Printing the results

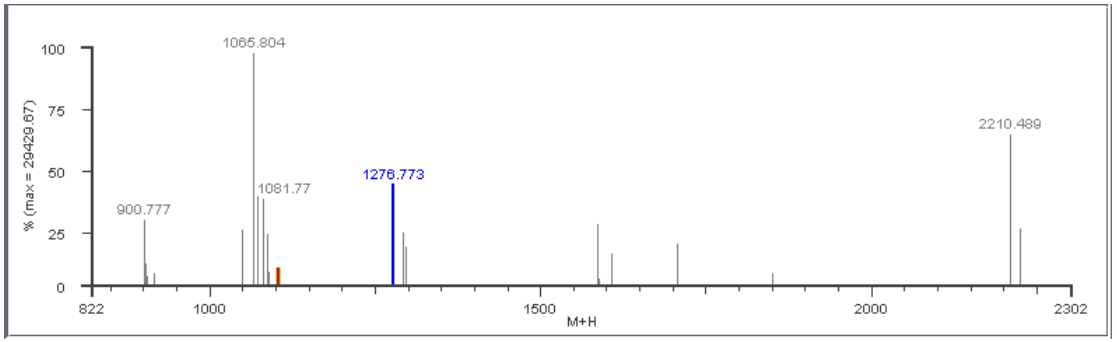
To print a summary of the workflow results, right-click either the protein or peptide table, and then click Print Workflow. Printing is controlled using the Print wizard (see Using print wizards on page 11-3).

Spectrum Viewer for MS data

For a search with an MS spectrum, the bottom component of the results browser is a graphical display of the MS spectrum data used for the search.

For a search with an MSMS spectrum, the middle component of the results browser is a graphical display of the parent spectrum from the MSMS spectrum data used for the search.

Spectrum Viewer for MS data:



In the graph:

X-axis = retention time

Rule: X-axis = mass if the spectrum data does not include retention times

Y-axis = intensity

Each peak is labeled with peak mass.

You cannot directly select results in the Spectrum Viewer. However, the viewer responds to selections in the other browser components and colors the peaks in the spectrum to indicate the type of peptide:

- If a protein or EST is selected, the peaks that have been matched to peptides belonging to the selected protein or EST are colored.
- If a mass is selected, the corresponding peak is colored.
- If a peptide is selected, the peak that is matched to the selected peptide is colored.

The colors in the graph are:

Gray	The peak is not matched to a peptide from the current protein or EST.
Blue	A standard peptide (that is, with no modifications or missed cleavage sites).
Red	A peptide that contains one or more missed cleavage sites.
Green	A peptide that contains one or more post-translational modifications.

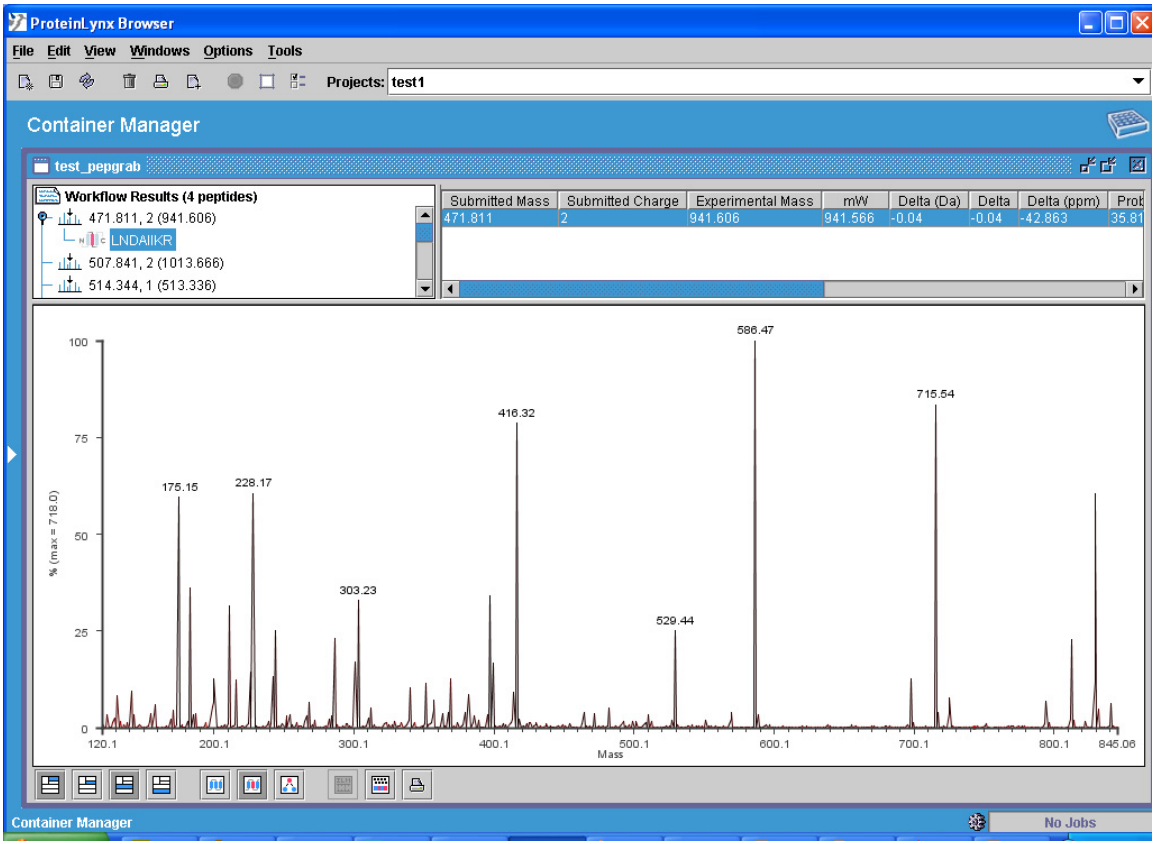
Yellow	A peptide that contains post-translational modifications and missed cleavage sites.

Viewing raw data

To view raw data, click the button to the right of the spectrum view. The processor needs to be running for the raw data to be retrieved as PLGS needs a live link to the raw data.

Result: A two-dimensional representation of mass (X-axis) against intensity (Y-axis) is displayed for the currently selected mass or peptide.

Raw data display:



There can be short delay between selecting the peptide in the tree and rendering the data for display.

In the graph, the coloring is:

Black = a high density of data.

Red = a low density of data.

To zoom into the raw data, use the zoom function, which is described in Spectrum Viewer options on page 6-24. As you zoom in to levels nearing that of the data, dots represent the actual mass intensity points. The graph color changes to red, which shows that the data is not dense.

Error messages

There are several error messages which could be displayed if there are problems retrieving the data; these are detailed in the following table.

Viewing raw data - error messages:

Error Message	Suggested Course of Action
Error connecting to processor, please start the processor	The raw data viewer needs the processor to be running; restart the processor. For details on starting the processor, see Chapter 1 - Installing ProteinLynx Global SERVER.
The raw data file requested was not found	The raw data viewer needs the original raw data file to be present. Ensure that the raw data file has not been deleted or moved since processing.
Invalid spectrum format, please re-process data	This indicates the spectrum is in an old format. Process the raw data again to update the spectrum.
The data requested was unavailable, please try again	This means the processor is running out of memory and has cleared the data. Try to reselect the node in the tree; if that does not work, restart the processor.

Viewing raw data - error messages: (Continued)

Error Message	Suggested Course of Action
The processor experienced an internal error. Please examine processor output.	This is an internal error and should be reported to Waters. Attach either the log file (see Chapter 1 - Installing ProteinLynx Global SERVER for assistance on locating the file) or a screenshot of the processor window (Ctrl+Print Scrn) to an e-mail, and send it to your local Waters support representative.
The processor did not accept the request	This is an internal error and should be reported to Waters.
Request parameters were invalid, no data was available	This is an internal error and should be reported to Waters.

Changing the x-axis view

If the mass spectrum data contains peak retention times as well as masses, you can choose to display either retention times or masses on the x-axis of the Spectrum Viewer.

To change the *x*-axis view, click

to show retention times on the *x*-axis, or

to show masses on the *x*-axis.

If retention times are displayed along the *x*-axis, the most intense peaks will be annotated with the peak mass.

If masses are displayed along the *x*-axis, the most intense peaks will be annotated with the peak retention time (or with the peak mass if the spectrum data does not include retention times).

Viewing the fragment ion display

To view the fragment ion display, click the button to the right of the spectrum view.

The fragment ion display shows the expected masses of the fragment ions for the predicted peptide sequence and the related delta masses of the experimental value. Ions that are shown in gray are undetected ions in the spectrum, and therefore do not have corresponding delta masses.

The ions found are colored according to the type of ion, using the color scheme on the MSMS spectrum display:

Gray	The peak is not matched to a peptide from the current protein or EST.
Blue	A standard peptide (that is, with no modifications or missed cleavage sites).
Red	A peptide that contains one or more missed cleavage sites.
Green	A peptide that contains one or more post-translational modifications.
Yellow	A peptide that contains post-translational modifications and missed cleavage sites.

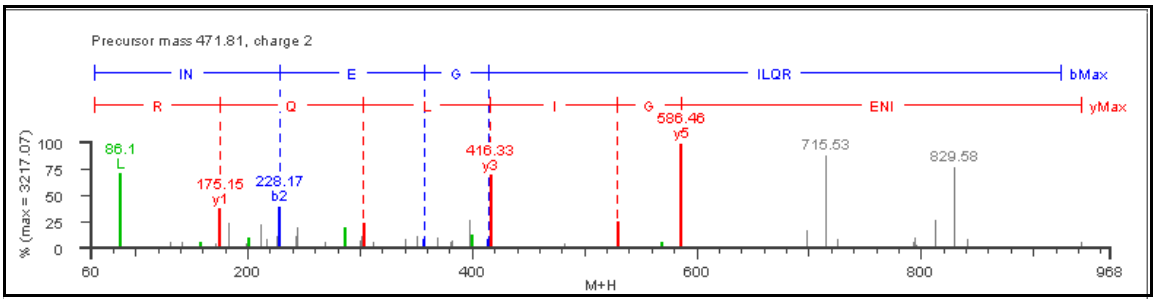
Fragment ion display for MSMS data:

a	60.045	159.113	272.197	359.229	430.267	559.309	672.393	769.446	883.489	996.573	1159.636	--
		(0.021)	(0.005)	(-0.014)		(0.031)	(0.017)	(0.044)	(-0.078)			
b	88.04	187.108	300.192	387.224	458.261	587.304	700.388	797.441	911.484	1024.568	1187.631	--
		(0.016)	(0.016)	(-0.003)		(-0.008)	(-0.087)	(0.04)				
	S	V	L	S	A	E	L	P	N	L	Y	R
	38.45	38.43	85.73	99.9	99.9	99.87	99.87	84.12	46.2	43.56	77.61	99.85
y"	--	1274.711	1175.642	1062.558	975.526	904.489	775.447	662.363	565.31	451.267	338.183	175.119
			(0.052)	(-0.012)	(-0.01)	(-0.029)	(0.02)	(-0.029)			(0.003)	(-0.001)
z	--	1257.684	1158.616	1045.532	958.5	887.463	758.42	645.336	548.283	434.24	321.156	158.093
		(0.081)		(-0.013)		(-0.04)	(-0.01)	(-0.023)			(0.005)	

Spectrum Viewer for MSMS data

For a search with an MSMS spectrum, the bottom component of the results browser is a graphical display of the fragmentation spectrum for the current parent peak.

Spectrum Viewer for MSMS data:



You cannot directly select results in the Spectrum Viewer. However, the viewer responds to selections in the other browser components:

- If a mass is selected, the fragmentation spectrum for the corresponding peak is displayed.
- If a peptide is selected, the fragmentation spectrum for the peak that is matched to the peptide is displayed. The graph is annotated with the fragmentation data for the peptide.

Peptide fragment annotation indicates the peaks that correspond to fragment ions from the peptide, and marks the positions of these ions within the peptide sequence.

The colors in the graph are:

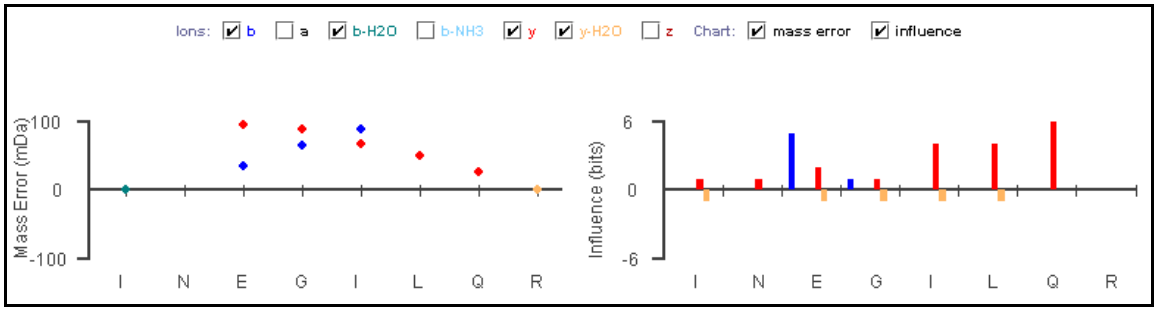
Red	y-series ions.
Blue	b-series ions.
Green	All other ions.

Displaying ion probabilities

To display ion probability data for the fragmentation spectrum, click the button to the right of the spectrum view.

To view data for one or more ion series, select each relevant check box on the display.

MSMS spectrum ion probabilities:



For each matched fragment ion, you can view either mass error or influence, or both:

- Mass error is the difference between the theoretical mass of a fragment ion and the peak mass from the spectrum to which the ion was matched. The peptide sequence is shown along the bottom of the graph, and each ion is indicated by a colored dot above the relevant position in the sequence. The color of the dot indicates to which series the ion belongs. The vertical position of the dot indicates the mass error.

  To view the mass error for the selected ion series, select the check box labeled mass error. To hide the mass error data, clear the check box.

  Rule: At least one of the graphs must be displayed at all times – if the influence check box was already cleared, it will be reselected automatically.

- Influence indicates whether the prediction of the selected ion is having a positive or negative effect on the peptide score; the more positive the number, the more influential the prediction.

  The peptide sequence is shown along the bottom of the graph, and each ion is indicated by a colored bar above the relevant position in the sequence. The color of the bar indicates to which series the ion belongs. The height of the bar indicates the influence.

  To view the influence for the selected ion series, select the influence check box. To hide the influence data, clear the check box.

  Rule: At least one of the graphs must be displayed at all times – if the mass error check box was already cleared, it will be selected automatically.

To return to the MSMS spectrum view, click the button to the right of the spectrum view.

Spectrum Viewer options

Several Spectrum Viewer functions are the same, regardless of whether MS or MSMS data is being displayed:

- Viewing a selected X-axis range.
- Scrolling along the X-axis.
- Displaying a zoomed section of the graph in a separate window.

Viewing a selected x-axis range

Rule: This function is not available when viewing ion probability data.

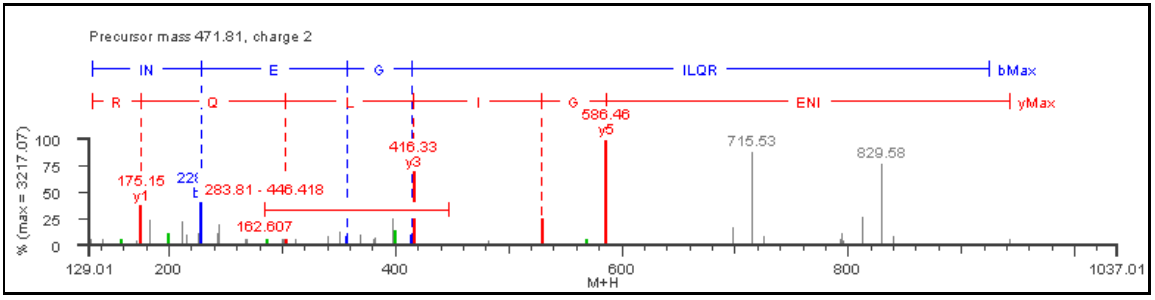
You can zoom in to a specific range along the x-axis.

To view an x-axis range:

1. Click and drag to select a range along the x-axis.

   A red line marks the selected range, which is labeled with the maximum and minimum X values in the range, and the length of the range. The selected range can be adjusted as long as the mouse button is held down.

Zooming in to a spectrum:



2. Release the mouse button. The X-axis range of the spectrum graph is altered to the selected range.

Repeat this procedure as often as needed. However, the length of the range must be at least 0.001 Da.

3. To zoom out again, either:
   - Right-click the Spectrum Viewer once to return to the previous range.
   - Right-click the Spectrum Viewer twice to return to the initial range (the full spectrum).

Scrolling along the x-axis

Rule: This function is not available when viewing ion probability data.

To scroll the graph, right-click and drag along the x-axis.

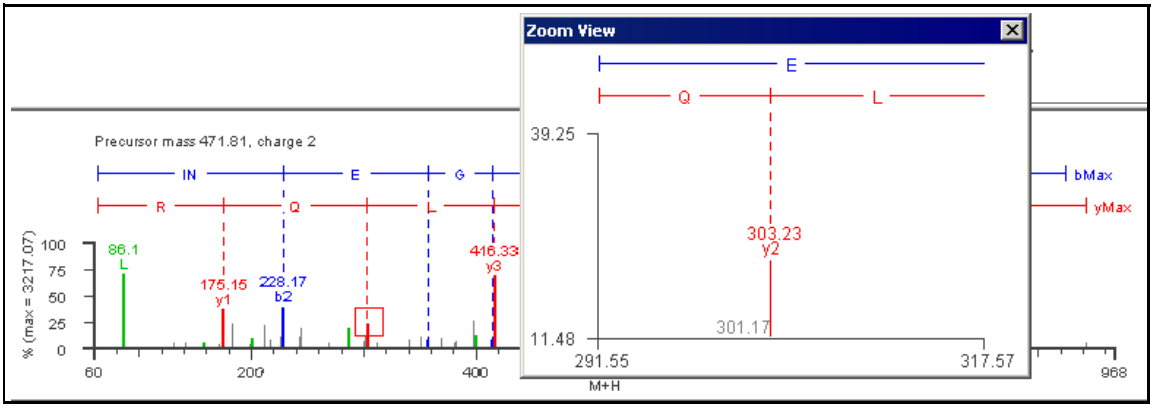
Displaying a zoomed section of the graph in a separate window

Rule: This function is not available when viewing ion probability data.

To display a close-up of a selected region of the graph in a separate window:

1. Double-click the Spectrum Viewer. A red box on the graph indicates the selected region. A separate window displays a close-up of the selected region.

Zoom View:



To alter the size and position of the selected region:

- To alter the size of the selected region, click on an edge of the red box and drag to adjust the size of the box.

- To select a different region, click inside the red box and drag to move to a different region.

Tip: The close-up window updates automatically as the size or position of the selected region is adjusted.

2. To close the separate window and remove the red box from the main graph, click the X in the top right corner of the separate window.

Copying data

To copy the spectrum data or ion probabilities data, click the button to the right of the spectrum view.

Copying spectrum data

If the spectrum viewer is showing a graph of the spectrum data, the data on the clipboard is arranged to show a paired X-value and Y-value on each line. The format is:

<X-value> <Y-value>

Copying ion probabilities data

If the Spectrum Viewer is showing ion probabilities, a list of mass errors and influences is copied to the clipboard for each ion series that is being displayed.

The top line of the copied data shows the name of each ion series, separated by a space. Each subsequent line shows an amino acid from the peptide sequence, followed by:

- the mass error for the first selected ion series
- the influence for the first selected ion series
- the mass error for the second selected ion series, and so on

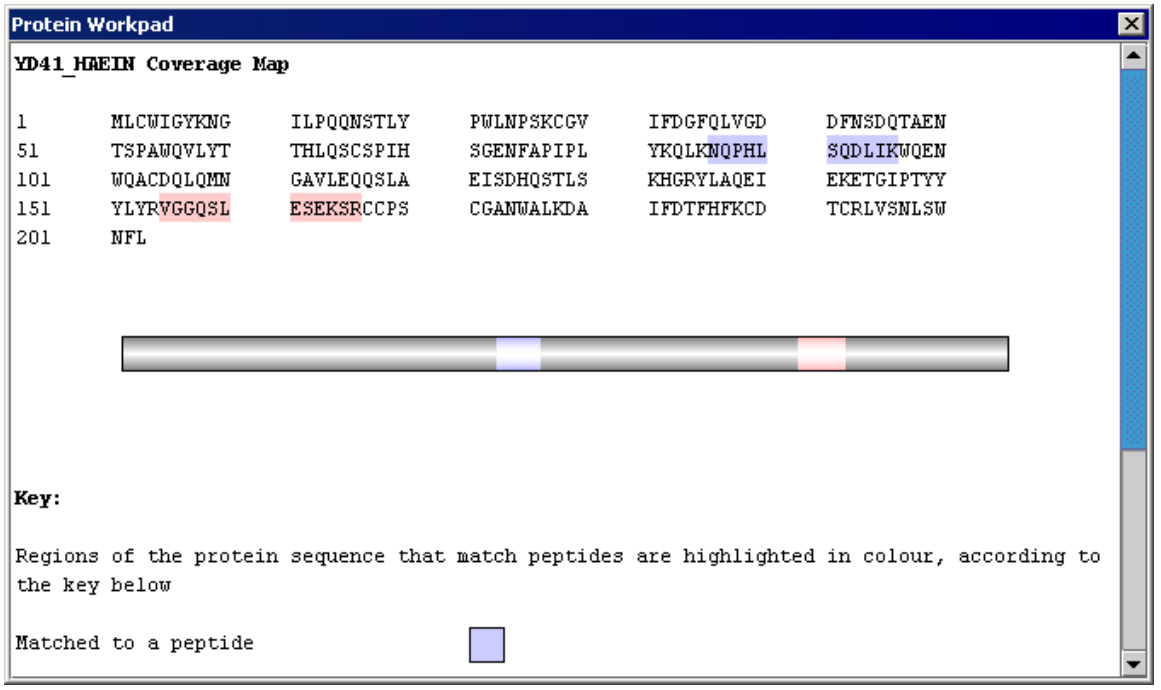
Each entry is separated by a space.

Protein Workpad

The Protein Workpad is a separate window that displays details of the currently selected protein or EST.

To view the Protein Workpad, right-click either the protein or EST table, and then click Protein Workpad.

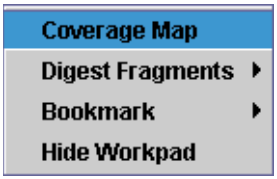
Protein Workpad:


Protein Workpad
YD41_HAEIN Coverage Map
1 MLCWIGYKNG ILPQQNSTLY PWLNPSKCGV IFDGFQLVGD DFNSDQTAEN
51 TSPAWQVLYT THLQSCSPIH SGENFAPIPL YKQLKNQPHL SQDLIKWQEN
101 WQACDQLQMN GAVLEQQSLA EISDHQSTLS KHGRYLAQEI EKETGIPTYY
151 YLYRVGGQSL ESEKSRCCPS CGANWALKDA IFDTFHFKCD TCRLVSNLSW
201 NFL
Key:
Regions of the protein sequence that match peptides are highlighted in colour, according to the key below
Matched to a peptide

Initially, the protein workpad shows a coverage map of the currently-selected protein or EST (see Coverage map on page 6-28).

To change the view, right-click in Protein Workpad. A pop-up menu opens.

Protein Workpad pop-up menu:



The menu items are:

- Coverage Map – shows the protein sequence and peptide matches.
- Digest Fragments – enables you to run simulated digests (see Running a simulated digest on page 6-29).
- Bookmark – enables you to retrieve the databank entry for the current protein or EST (see Retrieving databank entries on page 6-30).
- Hide Workpad – closes the Protein Workpad.

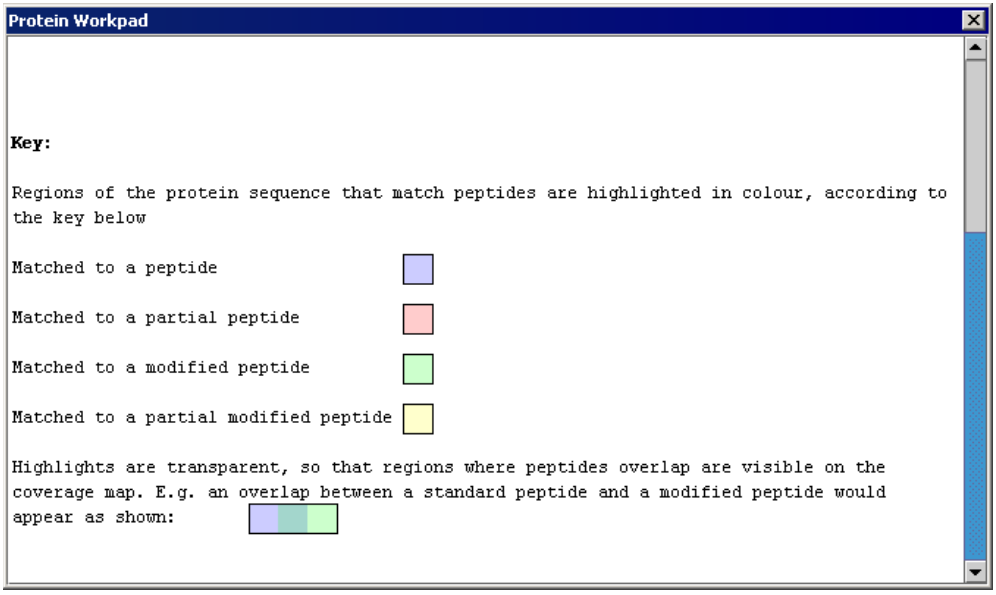
Coverage map

The coverage map shows the protein sequence and a graphical representation of the location of peptide matches.

The protein sequence is highlighted to indicate the location of peptide matches. The color of a highlight depends on the status of the peptide it represents. If several peptides cover a particular section of the sequence, this section will be a mixture of the highlight colors for the various peptides (if they are different in color), or a darker shade of the highlight color (if the highlights are the same color).

The highlight colors are explained by a key at the bottom of the coverage map.

Protein Workpad key:

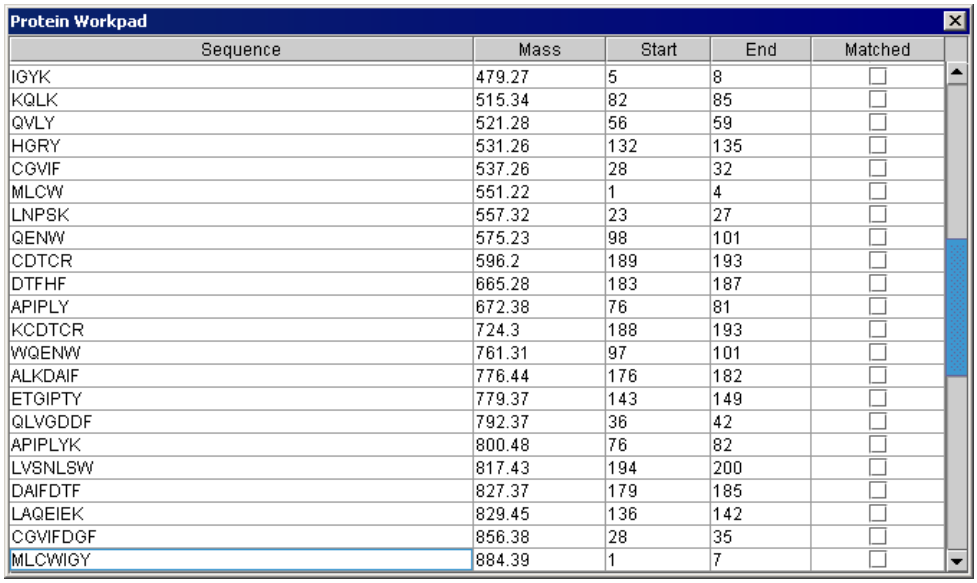

Protein Workpad
Key:
Regions of the protein sequence that match peptides are highlighted in colour, according to the key below
Matched to a peptide
Matched to a partial peptide
Matched to a modified peptide
Matched to a partial modified peptide
Highlights are transparent, so that regions where peptides overlap are visible on the coverage map. E.g. an overlap between a standard peptide and a modified peptide would appear as shown:

Running a simulated digest

To run a simulated digest of the current protein or EST, right-click the Protein Workpad, click Digest fragments, and then click a digest reagent from the list.

Result: A table is displayed, showing the fragments produced by the simulated digest.

Protein Workpad digest fragments:



Protein Workpad

Sequence	Mass	Start	End	Matched
IGYK	479.27	5	8	☐
KQLK	515.34	82	85	☐
QVLY	521.28	56	59	☐
HGRY	531.26	132	135	☐
CGVIF	537.26	28	32	☐
MLCW	551.22	1	4	☐
LNPSK	557.32	23	27	☐
QENW	575.23	98	101	☐
CDTCR	596.2	189	193	☐
DTFHF	665.28	183	187	☐
APIPLY	672.38	76	81	☐
KCDTCR	724.3	188	193	☐
WQENW	761.31	97	101	☐
ALKDAIF	776.44	176	182	☐
ETGIPTY	779.37	143	149	☐
QLVGDDF	792.37	36	42	☐
APIPLYK	800.48	76	82	☐
LVSNLSW	817.43	194	200	☐
DAIFDTF	827.37	179	185	☐
LAQEIEK	829.45	136	142	☐
CGVIFDGF	856.38	28	35	☐
MLCWIGY	884.39	1	7	☐

Retrieving databank entries

Use a bookmarked sequence databank search tool to retrieve the databank entry for the current protein or EST.

To carry out the search, right-click the Protein Workpad, click Bookmark, and then choose the Web-based sequence-retrieval system to use.

The results of the search are displayed in a browser window.

To add more sites to the bookmarked list, use the Bookmarks tab in the ProteinLynx Browser Preferences dialog box (see Bookmarks tab on page 2-11).

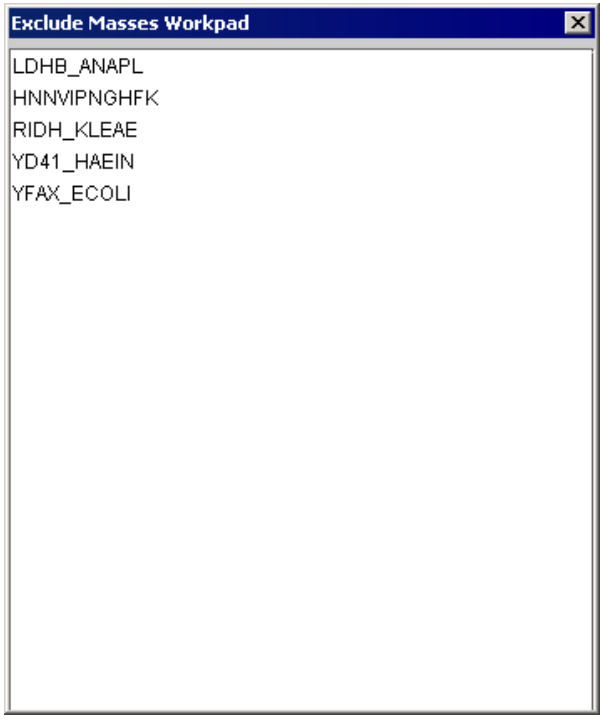
Exclude Masses Workpad

The Exclude Masses Workpad is a separate window that displays a list of items to exclude from any resubmitted searches using the current workflow.

Note: Masses selected for exclusion are usually theoretical masses, which can differ from masses found in the data. Therefore, due to the possibility of misassignment (a detected mass being mistaken for a different theoretical mass), the corresponding data is suppressed according to how well the masses match the theoretical masses rather than being completely extinguished.

To open the Exclude Masses Workpad, right-click either table, and then click Exclude/Re-submit > Open Exclude Mass Pad.

Exclude Masses Workpad:



For other options in the Exclude Masses Workpad, right-click the workpad to display the menu.

The menu items are:

- Add Exclude – There are four ways to add items to the Excluded Masses Workpad (see Adding items to the excluded list on page 6-32).
- Delete Exclude – Delete item from the Excluded list (see Deleting items from the excluded list on page 6-33).
- Use Reagent – Add an item that represents a digested protein or EST to the Excluded list (see Running a simulated digest for a protein on page 6-33).
- View Exclude Masses – View the mass values associated with an item (see Viewing the masses associated with an excluded item on page 6-34).
- View Protein Workpad – Open the Protein Workpad (see Protein Workpad on page 6-27).
- Hide Workpad – Close the Protein Workpad.

Adding items to the excluded list

There are five ways to add items (masses, proteins, and peptides) to the Excluded Masses Workpad:

- To add a mass shown in the peptide tree:
  1. In the workflow results window, click the Show Peptides/masses button, .
  2. From the navigation tree, drag the mass you wish to add onto the Exclude Masses Workpad.
- To add a protein shown in the protein tree:
  1. In the workflow results window, click the Show Proteins button, .
  2. From the navigation tree, drag the protein you wish to add onto the Exclude Masses Workpad.
- To add a peptide shown in the protein tree:
  1. In the workflow results window, click the Show Proteins button, .

2. Expand the navigator tree to show the peptides you want to exclude.
3. From the tree, drag the peptide you wish to add onto the Exclude Masses Workpad.

- To add a single mass value to the list:
  1. Right-click the Exclude Masses Workpad.
  2. Click Add Exclude > Add Mass.
  3. Type a mass value in the Add Exclude Mass dialog box, and then click OK.
- To add a common compound:
  1. Right-click the Exclude Masses Workpad.
  2. Click Add Exclude > Add From Library.
  3. Click the desired item in the drop down list, and then click OK.

Deleting items from the excluded list

To delete an item from the Excluded Masses Workpad:

1. Click the item in the Excluded Masses Workpad.
2. Right-click, and then click Delete Exclude.

Tip: To select multiple items, press Shift or Ctrl while clicking.

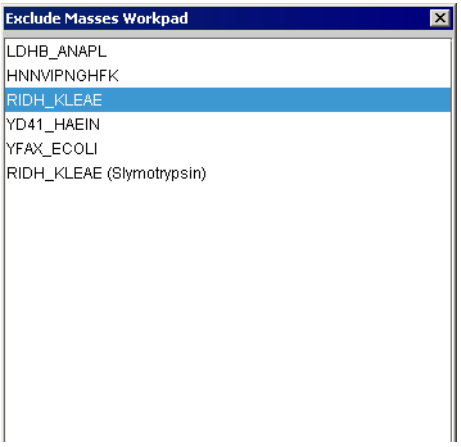
Running a simulated digest for a protein

To add a new item that represents a digested protein or EST to the Exclude Masses Workpad:

1. Click a protein or EST in the Excluded Masses Workpad, and then right-click.
2. Click Use Reagent.
3. Click a digest reagent in the list.

Result: A new item representing the digested protein or EST is added to the list.

Exclude Masses Workpad with digested protein added:



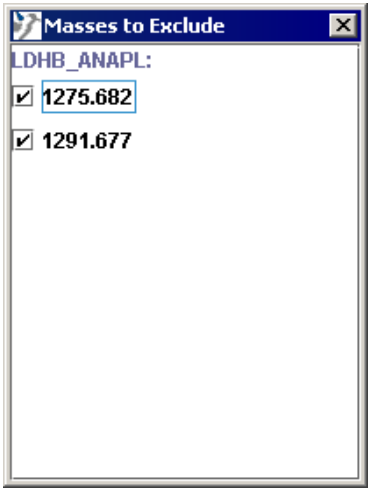
Viewing the masses associated with an excluded item

To view the mass values associated with an item in the Exclude Masses Workpad:

1. Click an item in the Exclude Masses Workpad, and then right-click.
2. Click View Exclude Masses.

   Result: A separate window is displayed, showing a list of the mass values.

Masses to Exclude window:



3. Select a check box to exclude that specific mass from resubmitted searches using the current workflow.
    - Items that represent an individual mass (that is, a mass entered by the user or a single peak mass from the spectrum) have only one associated mass - the mass value.
    - Items that represent a peptide have only one associated mass - the molecular weight of the peptide.
    - Items that represent a hit, protein, or EST have multiple associated masses. Each associated mass is the molecular weight of a peptide that is a match to the protein or translated EST sequence.
    - Items that represent a digested protein or EST have multiple associated masses that represent the molecular weights of peptides, but in this case the peptides are the fragments produced by the simulated protein digest.

7 Defining templates for searching with Workflow Designer

The Workflow Designer enables you to define a template that can be used to perform an automated databank search of samples in the Container Manager and Gel Manager.

Contents:

Topic	Page
What is Workflow Designer?	7-2
Creating a workflow template	7-5
Filters	7-11

What is Workflow Designer?

The Workflow Designer enables you to define a template that can be used to perform an automated databank search of samples in the Container Manager and Gel Manager.

To search MSMS, MS, or PSD data, you can use these search types:

- PMF (Peptide Mass Fingerprint)
- PMF + Fragment Ion Search
- Fragment Ion Search

To search Expression (MS^E) data, use these search types:

- Electrospray-MS (for low energy MS^E only)
- Electrospray High/Low

For each of these, you can use the Databank Search Query search method to identify a set of protein sequences. However, if you use a Fragment Ion Search only, you can also link this method with other search methods. Doing so progressively filters the search and analyzes the data more accurately. These other search methods are:

- AutoMod Query
- De Novo Query
- BLAST (Basic Local Alignment Search Tool) Query

If these are used, the results of one search are filtered to form the query of the next. This can significantly increase the number of peptides matched to fragmentation spectra data, and improves the coverage of the ESTs or proteins in the results.

You can save the workflow templates for use in other sessions.

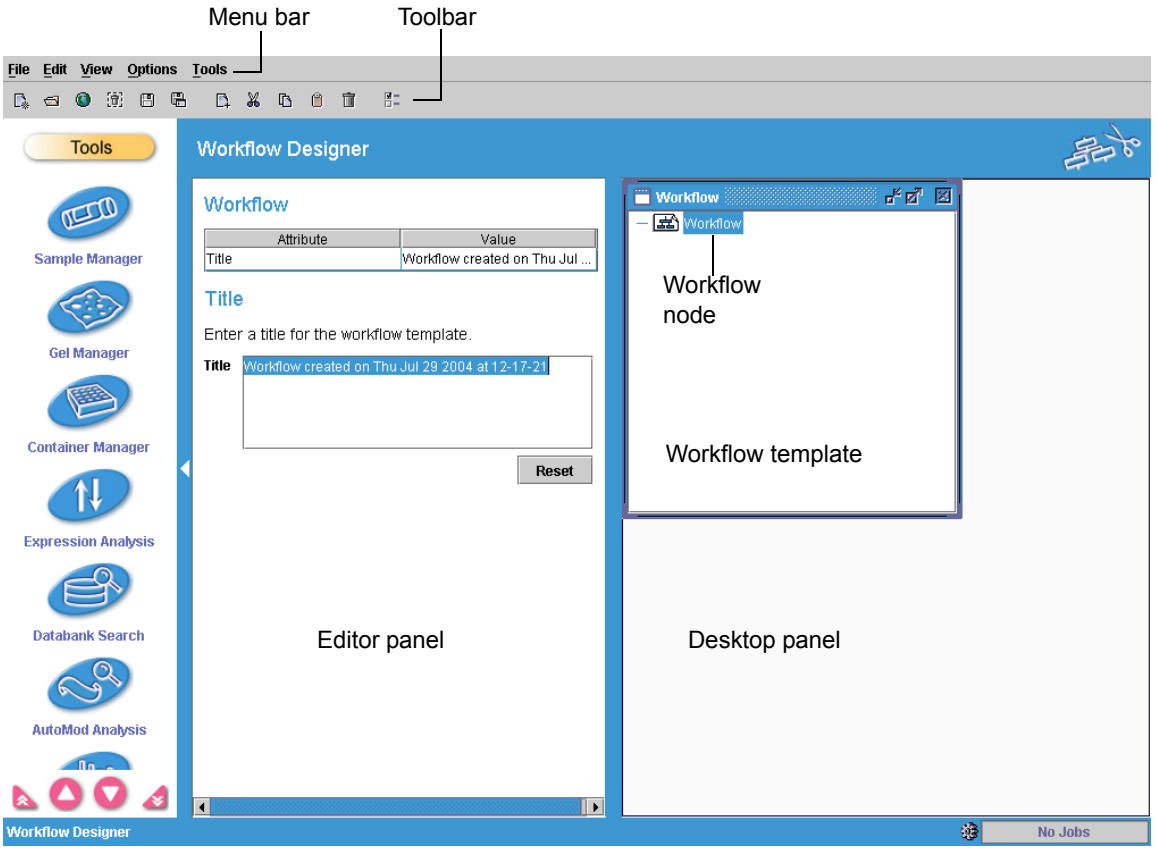
The Workflow Designer interface

To open the Workflow Designer, click the Workflow Designer icon in the tool tray.

The Workflow Designer opens with nothing displayed in the main window. When you have created a new template, the interface contains the following elements:

- Editor panel – Displays the attributes for the workflow and search methods.
- Desktop panel – Displays workflow templates.
- Workflow Template – Displays the search methods to be used for a workflow.
- Workflow node – Enables you to attach search methods to create a search strategy.

Workflow Designer - new template:



Workflow Designer toolbar

The following table describes the buttons on the Workflow Designer toolbar, and their corresponding menu bar options.

Workflow Designer toolbar options:

Button	Menu Bar Option	Description
	File > New	Adds a new workflow template to the desktop panel.
	File > Open	Opens a previously saved workflow template.
	File > Open URL	Opens the URL chooser dialog box (see Figure titled "URL Chooser dialog box:" on page 7-10) to enable you to specify a remote source that contains a workflow template.
	File > Remove	Removes the selected workflow template internal frame and discards all changes.
	File > Save	Saves the selected template.
	File > Save As	Prompts for a name and saves the selected template.
	File > Print	Prints the workflow template and all its automation parameters.
	Edit > Add	Opens a list which shows all the available automation tasks that can be added to the template.
	Edit >Cut	Removes the selected node and all of its children and stores them for use in the paste operation.
	Edit > Copy	Copies the selected node and all its children.
	Edit > Paste	Attaches the previously copied/cut node hierarchy to the position selected.
	Edit > Delete	Deletes the currently selected workflow node and all its children.
	Options > Preferences	General ProteinLynx preferences.

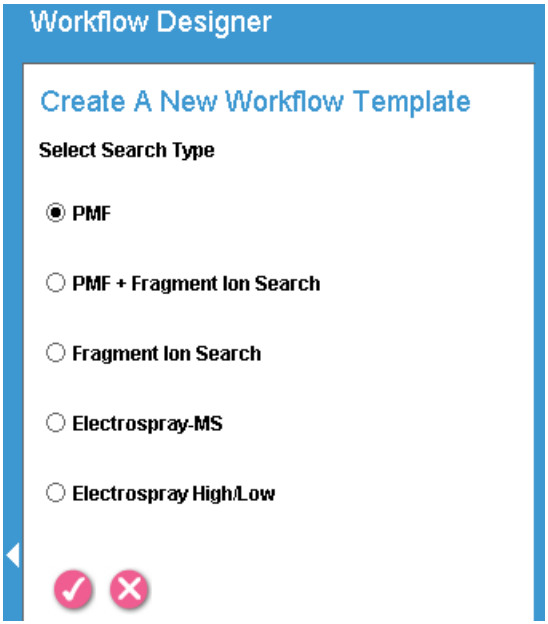
Creating a workflow template

To create a new workflow template:

1. Click on the toolbar.

   A panel is displayed, which enables you to select a search type for the template.

   Workflow Designer - selecting a type of search:



2. Select a search type, and then click .

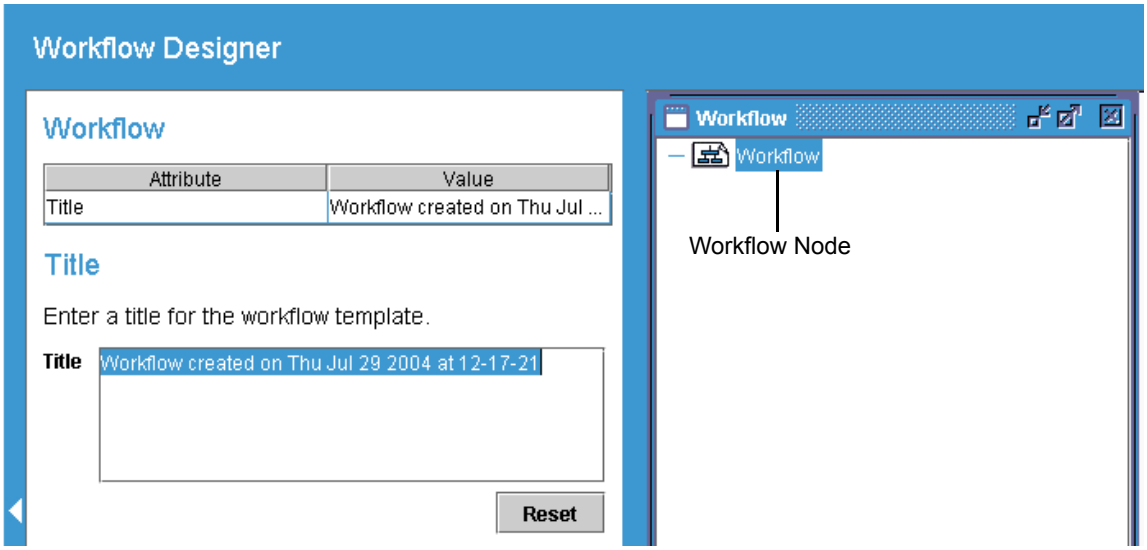
   Tips:

   - Fragment Ion Searches can be performed from any instrument that can generate fragmentation spectra. Therefore, Fragment Ion Searches can be performed on Electrospray Q-Tof, Maldi PSD and Maldi Q-Tof data.
   - The Electrospray-MS option enables searching of low energy MS^E data only; effectively a peptide mass fingerprint.
   - The Electrospray High/Low option enables searching of both the low and high energy MS^E fragment data.

Result: A new workflow template containing a new workflow node is displayed. You will attach search methods (queries) to this node.

By default, the title of the template is the current date and time, which is shown in the Editor panel. If desired, type a new title in the Title text box.

Workflow Designer - workflow node:

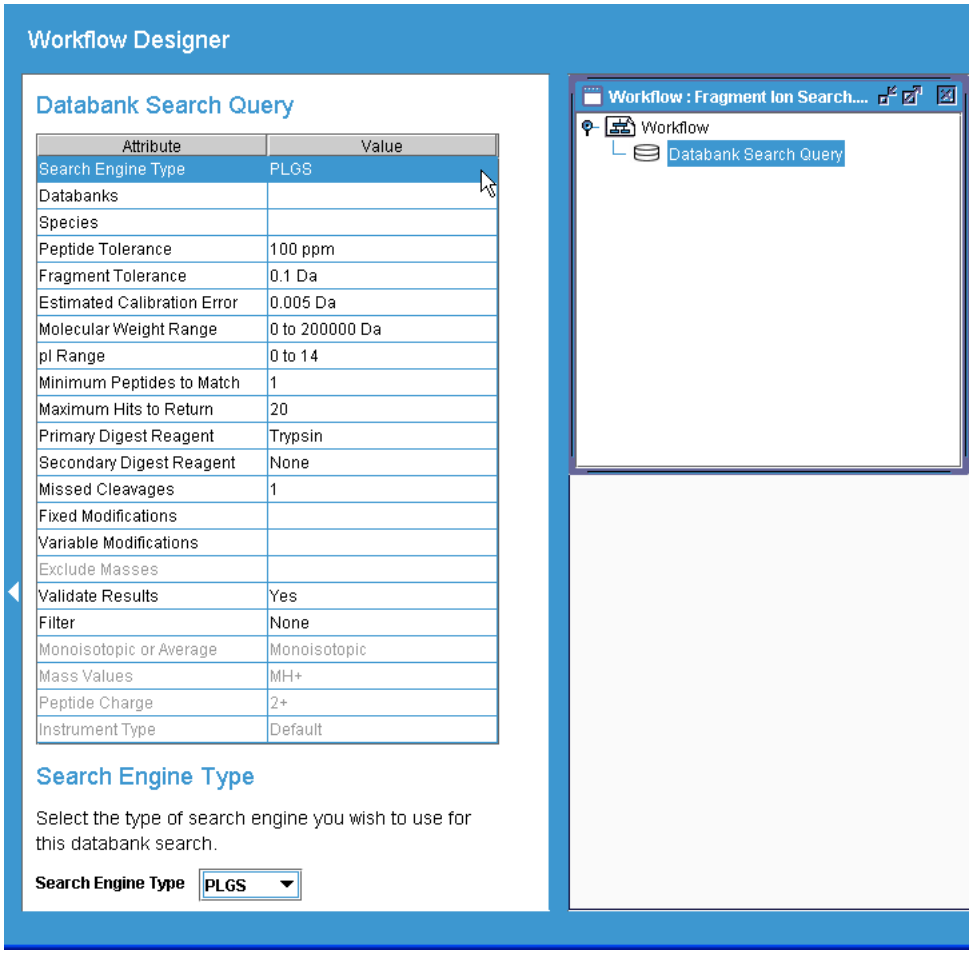


3. Right-click the workflow node, and then click Add.
4. If this is the first time that you have attached a search method, click Databank Search. For some search types, Databank Search is the only available option.

   The attributes displayed in the attribute table of the Editor panel vary slightly depending on the type of search engine: PLGS or MASCOT.

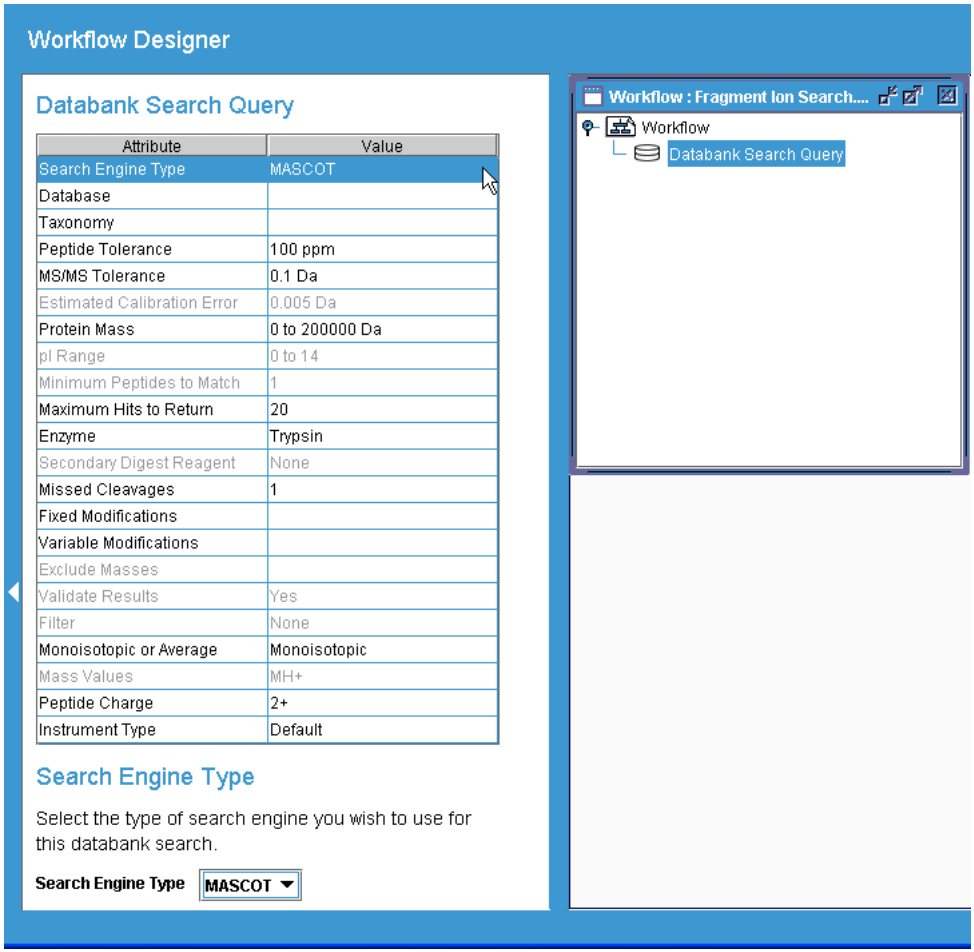
   Rule: MASCOT is only available for selection if you specify a Mascot search engine in the browser Preferences dialog box. See Search Engine tab on page 2-5.

Databank Search attributes - PLGS search engine:



For details of these attributes, see Databank search parameters on page 14-5.

Databank Search attributes - Mascot search engine:



For details of these attributes, see Databank search parameters on page 14-5.

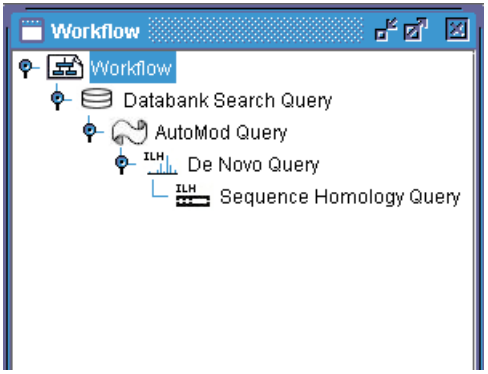
5. Set the attributes for the search as required.
6. If you want to add other search methods for a Fragment Ion Search, the following sequence is suggested:
    1. Databank Search — To identify a set of protein sequences to be analyzed further (see Databank Search tool on page 14-3).

2. AutoMod Query — To characterize the protein sequences fully by considering non-specific cleavages, amino acid modifications and substitutions (see AutoMod Analysis tool on page 14-14).
3. De Novo Query — To resubmit any fragmentation data that fails to match a peptide (see De Novo Sequencing tool on page 14-19).
4. BLAST (Sequence Homology) Query — To search novel peptide sequences against a databank to provide matches to homologous proteins (see BLAST Searching tool on page 14-23).

The attributes and values for each method are displayed in the Editor panel.

Tip: The selected search method is added directly to the node highlighted. For example, to add an AutoMod Query to a Databank Search Query, the Databank Search Query node must be highlighted, not the workflow node.

Typical workflow:



To reset the template name at any time before saving the template, click the workflow node, and then click Reset. This clears the Title text box and the value of the Title attribute. You can then type the new title in the Title text box.

7. To save the template, click on the toolbar.

Editing workflow templates

Workflow templates can be edited using the cut, copy, paste, and delete options available by right-clicking in the workflow template panel, or by using standard Windows keyboard shortcuts.

Rule: Editing the last search method on a branch will edit only that method. However, editing any other search method will affect all the results returned below it.

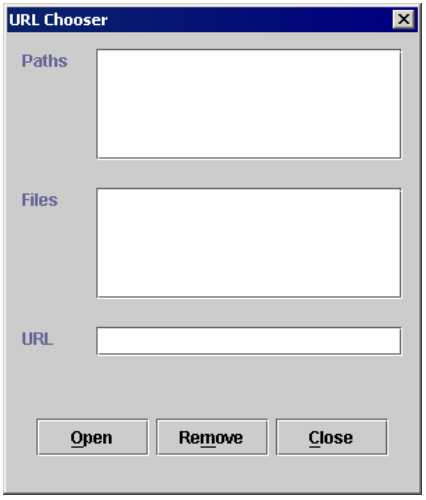
Opening workflow templates

Workflow templates are saved as XML (*.xml) files, and can be opened either from folders or from a URL.

To open a URL:

1. Click File > Open URL, or click on the toolbar.

   URL Chooser dialog box:

2. Specify the address in the URL field.
3. Click Open.

A list of previously opened templates will be listed in the Paths and Files fields each time the dialog box is reopened.

Filters

When several searches are chained together, the results of one search are filtered before being submitted as a query by the next search. For Databank Searching, AutoMod Analysis, and De Novo Sequencing, you can define this filtering process by specifying an XSL (eXtensible Stylesheet Language) style sheet.

XSL is a World Wide Web Consortium (W3C) standard defining style sheets for (and in) eXtensible Markup Language (XML) files.

- The XSL style sheet for a particular search tool is required to define which of the results that it receives from a prior search will be used to formulate its query.
- Default filters for AutoMod analysis (AutoMod_filter.xsl) and De Novo sequencing (DeNovo_filter.xsl) are provided. These two filters are sufficient for the majority of workflow templates.

AutoMod filter

The default AutoMod filter discards proteins that have a score less than zero. Therefore, only proteins with scores above zero undergo a theoretical digest and subsequent modifications, substitutions, and deletions.

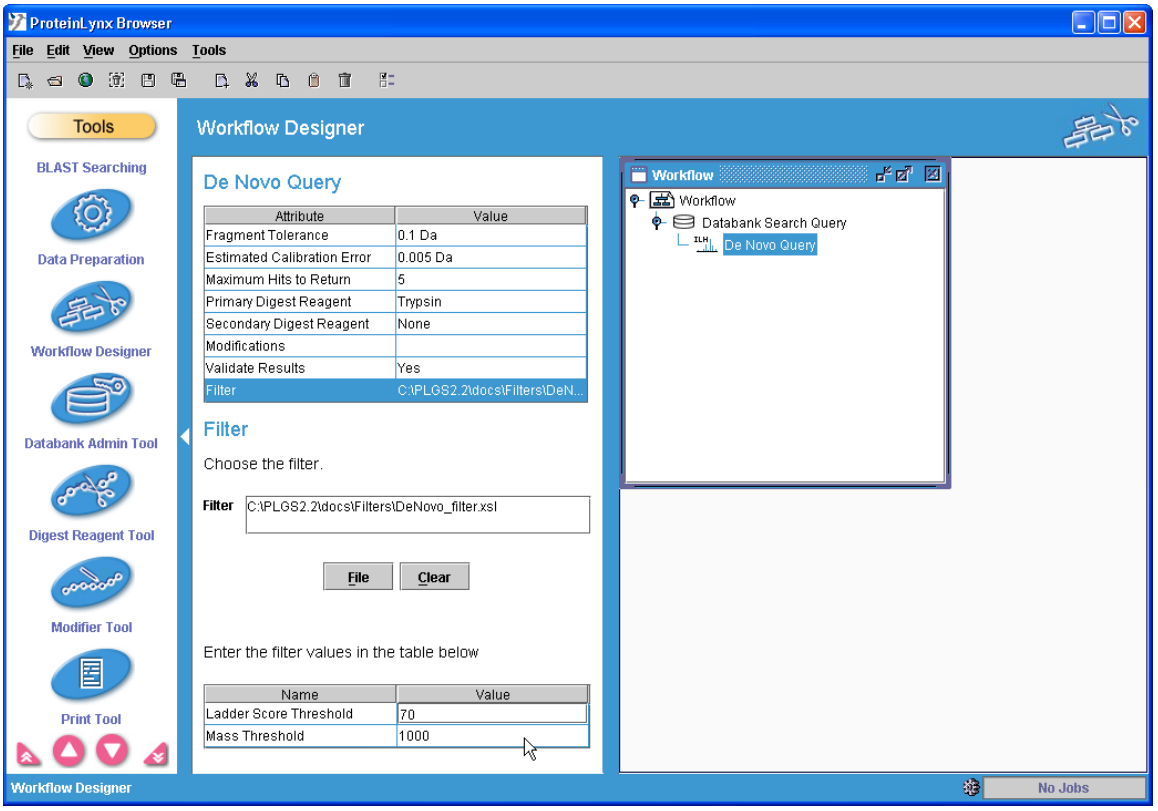
De Novo filter

The De Novo filter enables the default threshold values of different parameters to be altered through the browser, without having to modify the XSL document. The filter provided enables the ladder score and precursor mass thresholds to be amended.

The ladder score is based on the number of 'b' and 'y' ions in the peptide. The more b and y ions there are, the higher the score. The more consecutive b and y ions, the higher the score. y ions also contribute a greater score (up 66%) than b ions.

In the following example, only the MS/MS spectra of precursor masses greater than 1000 Da, that have not matched a peptide with a ladder score greater than 70, will be submitted for sequencing.

De Novo Query - Filter parameter:



The File button opens a file navigation dialog box, which enables you to select an XSL file. The XSL file specifies filter parameter names and values, which are displayed in the table of filter values.

The Clear button removes the reference to the XSL file and also the table of filter parameter names and values.

8 Creating custom processing parameters

The Data Preparation tool enables the creation of custom processing parameters, which are attached to raw spectra before processing.

Contents:

Topic	Page
Getting started with the Data Preparation tool	8-2
Attribute sets for data preparation	8-5

Getting started with the Data Preparation tool

Processing parameters templates determine how the RAW spectrum data is processed and whether certain attributes (for example, smoothing) are considered.

To open the Data Preparation tool and create a new template:

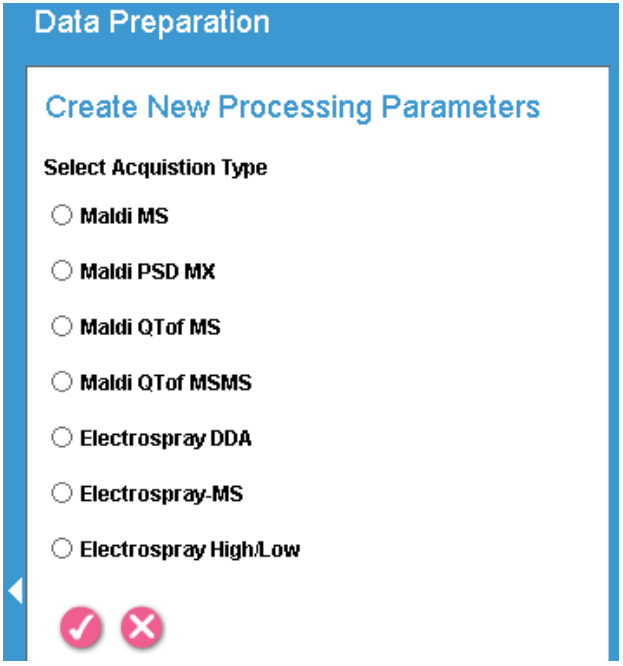
1. Click the Data Preparation icon on the tool tray.

   The Data Preparation window opens. Nothing is displayed in the main window.

2. Click on the toolbar.

   A panel appears, from which you can select an acquisition type for the template.

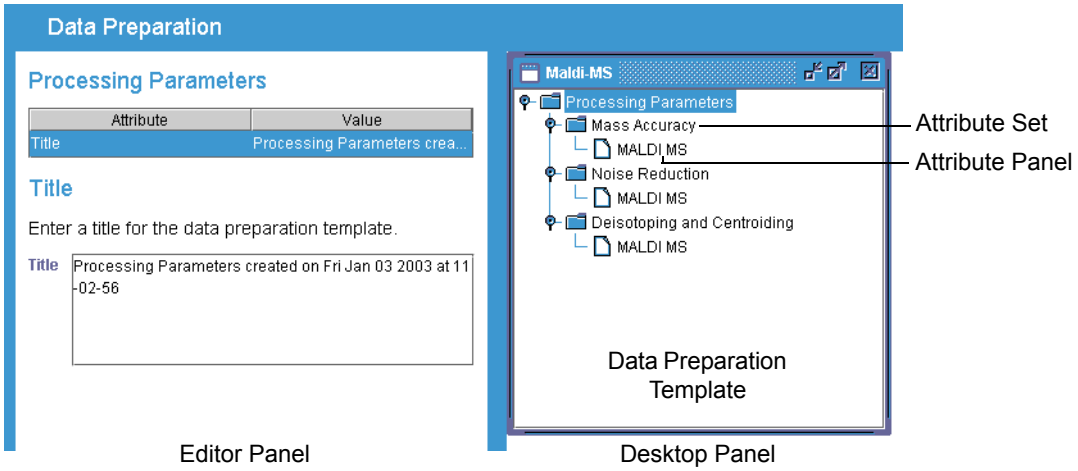
Data Preparation tool - selecting a type of acquisition:



3. Select the type of acquisition that generated the raw data, and then click .

   A data preparation template is displayed on the Desktop panel and an Editor panel is displayed in the left-hand panel. The next graphic shows a new MALDI-MS processing template.

Data Preparation tool display:



   By default the title of the template is the current date and time, which is shown in the Editor Panel. If desired, type a new title in the Title text box.

   The Data Preparation template for each acquisition type (instrument) has similar attribute sets and attribute panels. However, the attributes available in the attribute panels depends on the selected acquisition type.

   Click the relevant file icon in the template to display the attribute panel in the Editor Panel on the left of the screen. The details of each attribute are displayed under the attribute panel.

To save the processing parameters template, either:

- Click the Save button on the toolbar, or
- Click File > Save.

To remove the processing parameters template you are currently editing, either:

- Click the Remove button on the toolbar, or
- Click File > Remove.

  Note: If you are editing an existing template, the XML file will not be deleted; the displayed template frame and attribute list will just be cleared.

Attribute sets for data preparation

There are seven methods used to acquire data:

- MALDI MS
- MALDI PSD MX
- MALDI Q-Tof MS
- MALDI Q-Tof MSMS
- Electrospray DDA
- Electrospray-MS
- Electrospray High/Low

For each acquisition type, you can specify the following sets of attributes in the processing parameters templates:

- Mass Accuracy
- Noise Reduction
- Deisotoping and Centroiding
- Peak Matching – MALDI PSD MX only
- Chromatogram – Electrospray-MS and Electrospray High/Low only

Restriction: Some attributes in the attribute panels are disabled, and these cannot be edited. Some of these grayed-out attributes have default values that are used by the processor.

MALDI PSD MX

For the Noise Reduction and Deisotoping and Centroiding attributes, two template panels (MALDI MS, PSD MX) are displayed, which have related attributes. The panels labeled MALDI MS represent the processing to apply to MALDI MS data; the panels labeled PSD MX represent the processing to apply to PSD MX data.

MALDI Q-Tof MSMS

For the Noise Reduction and Deisotoping and Centroiding attributes, two template panels (MALDI Survey, MSMS) are displayed, which have related attributes. The panels labeled MALDI Survey represent the processing to apply to survey data; the panels labeled MSMS represent the processing to apply to MSMS data.

Electrospray DDA (QTOF-MSMS)

For each attribute, two template panels (Electrospray Survey, MSMS) are displayed, which have related attributes. The panels labeled Electrospray Survey in each attribute represent the processing to apply to survey data; the panels labeled MSMS represent the processing to apply to MSMS data.

Mass Accuracy attributes

Not all attributes are available for all panels: check the Applies to column in the table below to see whether the attribute listed relates to the panel you are configuring.

Mass Accuracy attributes:

Attribute	Applies to	Description
Select Calibration Type	MALDI MS MALDI Survey Electrospray-MS Low Energy High Energy	The type of calibration that should be performed. INTERNAL should be selected when the lock mass is present in the analyte (such as Trypsin autolysis products). EXTERNAL should be selected when the data contains dedicated lock mass (reference or 'near point') scans.
External Lock Mass	MALDI MS MALDI Survey	Enter the 'near point' or 'external' Lock Mass. If the Lock Mass is found in the data within the specified tolerance, a linear calibration correction will be applied to the data.

Mass Accuracy attributes: (Continued)

Attribute	Applies to	Description
Primary Internal Lock Mass	MALDI MS MALDI Survey	The primary internal Lock Mass. This could be the mass of a trypsin autolysis peptide or another known component of the sample. If the Lock Mass is found in the data within the specified tolerance, a linear calibration correction will be applied to the data. This correction replaces any external correction.
Secondary Internal Lock Mass	MALDI MS MALDI Survey	The secondary internal Lock Mass. This will be used if the primary internal Lock Mass is not found.
Lock Mass tolerance	All	The Lock Mass tolerance. If no peak is found within the tolerance, no correction will be applied.
Intensity Threshold	MALDI MS (MALDI PSD MX only)	The number to be used when locating the lockmass peak. De-isotoped peaks with intensities below this threshold will not be considered as potential lock masses. Set the units for this in the Threshold Type attribute.
Threshold Type	MALDI MS (MALDI PSD MX only)	Select how the Intensity Threshold attribute is expressed: %BPI – A percentage of the base peak intensity. Counts – A specific number for the threshold.

Mass Accuracy attributes: (Continued)

Attribute	Applies to	Description
Perform Lock Spray Calibration	Electrospray Survey	Enable or disable Lock Spray calibration. Enable for data acquired using an external Lock Spray interface.
Lock Spray Lock Mass	Electrospray Survey MSMS Electrospray-MS Low Energy High Energy	The expected position of the external lockspray peaks. **Example:** For a doubly charged species with molecular mass 1569.6696 Da, this is 785.8426 Da/e. The Electrospray Survey value (preferably doubly charged) will be used to correct survey data, and the MSMS value (preferably singly charged) will be used for fragmentation spectra. **Rule:** The same lockspray function is used for survey and MSMS. If only one lock spray ion is present, the same value can be entered in the survey and MSMS boxes.
Lock Spray Scans	Electrospray Survey MSMS Electrospray-MS Low Energy High Energy	The number of consecutive Lock Spray spectra which should be summed to determine the mass correction for each precursor.

Noise Reduction attributes

Not all attributes are available for all panels: check the Applies to column in the table below to see whether the attribute listed relates to the panel you are configuring.

Noise Reduction attributes:

Attribute	Applies to	Description
Background Subtract Type	All	Background subtraction removes slowly varying (low frequency) components from the data. This can improve the results of subsequent processing. Select from: None – No background subtraction is done. Normal – Normal background subtract removes smooth, slowly varying components from the data. Adaptive – Adaptive background subtraction additionally removes noise with a structure that repeats every nominal mass (roughly 1Da). Adaptive background subtraction can be particularly useful for low concentration MALDI data.
Background Threshold	All	The algorithm will aim to find a smooth function which lies above this percentage of data points. The value of the function in each channel is then subtracted from the data.

Noise Reduction attributes: (Continued)

Attribute	Applies to	Description
Background Polynomial	All	The order of the polynomial with which to fit the background. A value of 0 corresponds to a flat threshold and 1 is a sloping straight line. For typical data a value of around 5 will be sufficient.
Perform Smoothing	All	Whether to perform smoothing. Smoothing removes rapid variations in intensity, and can improve peak detection results.
Smoothing Type	All	The smoothing method to use. Savitzky-Golay smoothing preserves line width better than Mean smoothing.
Smoothing Iterations	All	The number of times that the smoothing should be performed.
Smoothing Window	All	The half width of the smoothing window in channels.
Combine Options	MALDI MS (not MALDI PSD MX) MALDI Survey	The method of combining scans. The reference (external lock mass) scans are never combined with sample scans. The setting of this attribute will affect whether other attributes are available. **Recommendation:** The recommended setting is All.
Scans to Combine	MALDI MS PSD MX MALDI Survey	The number of scans to combine. This option is only available when Combine Options is set to User-input.

Noise Reduction attributes: (Continued)

Attribute	Applies to	Description
Low Mass Threshold	MALDI MS PSD MX MALDI Survey	The low mass threshold. Only data above this threshold is used to determine which scans to combine. This option is only available when Combine Options is set to Auto-select.
Intensity Range	MALDI MS PSD MX MALDI Survey	The intensity range to consider. The intensity is specified as a percentage of the maximum possible without saturating the detector. Only spectra whose maximum intensity peak (above the mass threshold) lies within this range will be combined. This option is only available when Combine Options is set to Auto-select.
Peptide Filter	MSMS	Whether to perform background subtraction. Background subtraction removes slowly varying (low frequency) components from the data. This can improve the results of subsequent processing.

Deisotoping and Centroiding attributes

Not all attributes are available for all panels: check the Applies to column in the table below to see whether the attribute listed relates to the panel you are configuring.

Deisotoping and Centroiding attributes:

Attribute	Applies to	Description
Perform Deisotoping	MALDI MS Electrospray Survey MSMS MALDI Survey PSD MX	Whether to perform deisotoping. All three types of deisotoping simplify the data by replacing each ion cluster with a single mass measurement that represents the Carbon 12 peak (monoisotopic peak). Yes – The results are expressed on a singly charged scale. No – The spectra are peak detected only; all isotopes are preserved.
Deisotoping type	All	The type of deisotoping to perform: slower is more rigorous. The three different types of deisotoping are controlled by different parameters, which become available or unavailable depending on the deisotoping type selected. Use the slider bar to select slow, medium, or fast.
Iterations	All	The number of iterations.

Deisotoping and Centroiding attributes: (Continued)

Attribute	Applies to	Description
Threshold	All	The threshold is a percentage of the area of the most intense peak in the spectrum, and is used as a guide to break the spectrum into independent blocks. Breaking up the spectrum simplifies the deisotoping problem and speeds up the solution.
Centroid Top	All	The top percentage of each peak to use to determine its centroid. This option is only available if deisotoping is not selected.
Minimum Peak Width	All	The minimum peak width. Peaks having widths smaller than this number of channels will be removed or merged with adjacent peaks. This option is only available if deisotoping is not selected.
Automatic Thresholds	Electrospray Survey MSMS	When automatic thresholding is used, the deisotoping algorithm attempts to choose a sensible threshold for every spectrum that it is given. Although processing the data in this way should give reasonable results, experienced users might wish to set thresholds manually to reduce the number of ions reported or to attempt to improve sensitivity.

Deisotoping and Centroiding attributes: (Continued)

Attribute	Applies to	Description
TOF Resolution	Electrospray Survey MSMS PSD MX Electrospray-MS Low Energy High Energy	TOF resolution is m/z divided by full peak width at half maximum. Used together with the NP multiplier to correct for detector deadtime.
NP Multiplier	Electrospray Survey MSMS PSD MX Electrospray-MS Low Energy High Energy	This attribute is used together with TOF Resolution to correct for detector deadtime.
Minimum Charges to Report	Low Energy	The minimum charge state to report. Contributions to ions from charge states lower than this value will be removed. **Recommendation:** A setting of 2 is recommended to reject singly-charged noise.
Maximum Number of Charges	Low Energy High Energy	The maximum charge state to use in deisotoping. This should be set to the maximum charge state that is commonly observed in the data (to allow deisotoping to be performed correctly), but no higher. Increasing this value increases processing time.

Peak Matching attributes

The Peak Matching attributes are only available for PSD MX panels.

Peak Matching attributes:

Attribute	Description
Number of Precursors	The number of ions to submit for peak matching. The most intense ions in the spectrum are selected.
Fragment Intensity Threshold	The intensity (number of counts) above which fragment peaks are considered to be signal.
Precursor Matching Window	The percentage of the precursor mass for the tolerance of the precursor masses.
Fragment Matching Window	The tolerance, in parts per million (ppm) of the fragment masses.
Report Monoisotopic Fragment Masses	Selected (Yes) – Monoisotopic fragments are reported. Cleared (No) – Average fragment masses are reported.
Calibration File	**Default:** None. File – Opens the File Chooser dialog box. Navigate and choose file. The file path and name are displayed in the box. Clear – Selects None.

Chromatogram attributes

The Chromatogram attributes are available for the Electrospray-MS and Electrospray High/Low panels.

Chromatogram attributes:

Attribute	Description
Minimum Peak Width	The duration (in scans or time) for which the threshold criterion must be met for a peak to be reported.
Expected Peak Width	The expected peak duration (full width half maximum). This is used to help decide when ions start and stop eluting.

Chromatogram attributes: (Continued)

Attribute	Description
Peak Width Units	The unit by which peak width should be measured.
Automatic Thresholds	When automatic thresholding is used, the deisotoping algorithm attempts to choose a sensible threshold for every spectrum that it is given. Although processing the data in this way should give reasonable results, experienced users might wish to set thresholds manually to reduce the number of ions reported or to attempt to improve sensitivity.
Threshold	The total number of ions (not the height) that the first peak in an isotope cluster (usually referred to as the C12 peak) must possess for the threshold criterion to be exceeded in a single scan. **Tip:** To estimate this, centroid a typical scan (containing analyte) in MassLynx and look for this peak in a small but well defined isotope cluster. Increasing the threshold can dramatically speed up processing by reducing the apparent complexity of the data.
Select time range	Whether or not to limit (by scans or retention time) the range of data that should be processed.
Select start time	The retention time at which processing should start.
Select stop time	The retention time at which processing should stop.
Range Units	The units in which the Time Range is specified.

9 Viewing and processing gel data with Gel Manager

Gel Manager lets you view and process gel data, with clear sample tracking from gel to sequence identification.

Contents:

Topic	Page
Getting started with Gel Manager	9-2
Adding and importing data	9-3
Processing data	9-8
Viewing gel data	9-9

Getting started with Gel Manager

You can perform various operations with Gel Manager:

- Gels and cut lists (lists of gel spots) can be imported from a project or sample list into a project. This enables gel spots to be mapped onto plates and viewed in the Container Manager.
- Individual samples can be submitted to MassLynx for automated data acquisition and processing.
- Workflows can be attached to samples for automated Databank Searching, AutoMod Analysis, BLAST (Basic Local Alignment Search Tool) Searching, and De Novo Sequencing.

To open the Gel Manager, click the Gel Manager icon in the tool tray.

Adding and importing data

Initially a project needs to be created or opened. To create a project see Importing and viewing PLGS sample lists on page 5-3.

Adding a new gel without an image

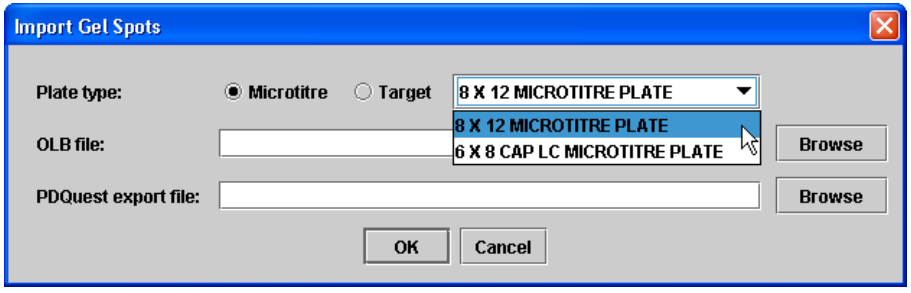
1. In the navigator tree click the Gels node, and then right-click.
2. Click Add Gel.
3. Type a name to associate with the gel in the ProteinLynx browser, and then click OK.

Importing gel spots

To import gel spots:

1. In the navigator tree, click the node of a gel you have created, and then right-click.
2. Click Import Gel Spots.

Import Gel Spots dialog box:



Import Gel Spots dialog box parameters:

Parameter	Description
Plate type	The Plate type onto which gel spots should be mapped. Also, select the specification of the plate from the drop-down list.
OLB file	A Waters-format OLB file that maps samples from the gel onto plates.

Import Gel Spots dialog box parameters: (Continued)

Parameter	Description
PDQuest export file	A PDQuest export file listing the co-ordinates of spots that were excised from the gel to create samples.

PDQuest files must be in plain text (.txt) or excel (.xls) format. OLB files must be in the Waters olb format (.olb).

3. Select the Plate Type.
4. Use the Browse buttons to select the relevant OLB and PDQuest files.

   Rule: Both the OLB file and the export file must be specified.
5. Click OK.

   The Specify Plates dialog box opens.

Specify Plates dialog box:

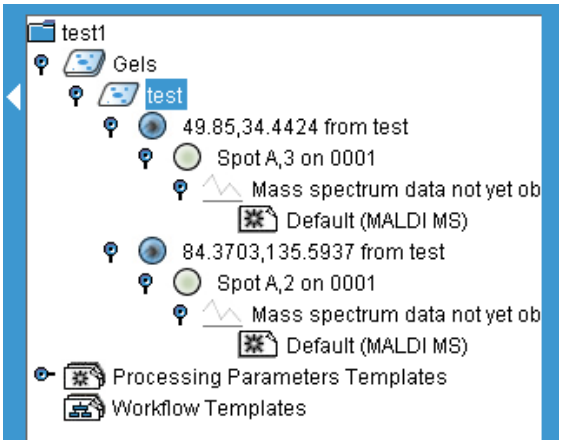
6. Select a plate from the ProteinLynx system or create a new plate record.

   Rule: If a new plate is created, a title or identifying number must be entered. If there is more than one plate listed in the OLB file then there will be a prompt for each plate.

Results:

- The specified plates are produced or updated as necessary in the Container Manager.
- When importing is complete, nodes are added beneath the gel node in the navigator tree to represent the imported gel spots.

Gel Manager navigator tree - gel data imported:



Further icons will be added to represent the plate wells or spots that the samples have been mapped to.

Importing a gel from an OLB file

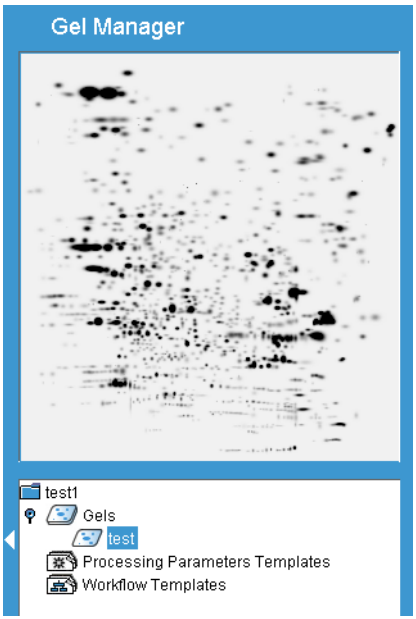
An OLB file is a system file of a gel image. This process only adds a gel image: you then have to associate OLB data.

To import gels from an OLB file:

1. In the navigator tree click the Gels node, and then right-click.
2. Click Import Gel.
3. Browse to the TIFF (*.tif) or JPEG (.jpg) gel image you wish to import. Click Open.
4. Type the name to associate with the gel in the ProteinLynx browser.

Result: When importing is complete, a new node is added to the navigator tree beneath the Gels node. Click the new node to display the gel image above the navigator tree.

Gel Manager navigator tree with gel Image:



Importing a gel from sample list

This process imports a gel image and gel spots.

To import gels from a sample list:

1. In the navigator tree click the Gels node, and then right-click.
2. Click Import Gel.
3. In the Files of Type list, click the type of sample list XML file you wish to import.
   - PDQuest XML file – The sample list XML file that can be exported from PDQuest software. The gel image, gel spot, container, and sample tracking information contained in the file are imported into the current project.
   - Progenesis XML file – The experiment XML file that can be exported from Progenesis Discovery software. The gel image and gel spot information contained in the file are imported into the current project.

As part of the import process, you must specify the plate names to which the gel spots will be mapped. As there is no sample tracking information in files of this type, gel spots are assigned to newly created containers in the order they are listed in the file.

Requirement: The gel image file must be in the same directory as the XML file selected.

4. Browse to the file, and then click Open.

Replacing the sample in a well or spot

To map a microtitre plate well or target plate spot to a different sample:

1. In the navigator tree, click the Well or Spot node, and then right-click.
2. Click Set Sample.

   Rule: The Set Sample option is not available if the current sample has been used to obtain mass spectrum data or workflow results.

Processing data

For details of the methods used for processing data, see:

- Chapter 2 - Setting up ProteinLynx Global SERVER – for details of attaching raw data files, workflow templates, and processed data.
- Chapter 7 - Defining templates for searching with Workflow Designer – for details of workflow templates.
- Chapter 8 - Creating custom processing parameters – for details of processing parameter templates.

Viewing gel data

Viewing a gel image

A gel image can be viewed by clicking the node of the gel in the navigator tree. The image can be manipulated in the following ways:

- If gel spots have been imported for the gel, the spots will be circled to mark their locations on the gel image.

  To remove these circles, right-click the image, and then clear the Circle Gel Spots check box.
- Right-click the image, and then select the Show Axis Labels check box. Labeled axes for the image are displayed.
- Zoom in to a region of the gel image – Select a region of the gel image by dragging a rectangle on the image. Zoom in to the selected region by double-clicking inside the rectangle.

  Repeat the procedure to zoom further into the image.

  To zoom out, double-click the image without selecting a rectangle first.
- Select a gel spot by double-clicking the gel spot on the image, or by selecting the gel spot icon in the navigator tree.

If workflow results have been obtained for the sample from the gel spot, the name of the top-scoring protein or EST from the search results is displayed when the mouse is moved over the gel spot.

Viewing a summary of results for a gel

Click the Gel icon in the navigator tree to view a gel summary. The summary tabulates the top-scoring protein or EST match for each spot in the gel. Each row includes the gel spot coordinates and similar information to that found in the corresponding workflow results windows (see Chapter 6 - Viewing results in the Results Browser).

Viewing sample annotation

To view the annotation for a sample in any given microtitre plate well or target plate spot, click the well or spot icon, right-click, and then click View Sample Information.

A sample display pane and results window are shown in the desktop area.

10 Using Expression Analysis to compare and analyze sample groups

Expression Analysis identifies and extracts pairs of labeled masses, computes their relative abundance, and indicates whether they are upregulated or downregulated. Expression Analysis enables you to perform expression profiling experiments.

Contents:

Topic	Page
Getting started with Expression Analysis	10-2
Experiment Analysis Design Manager	10-3
Viewing Expression Results	10-10
Log Plot Viewer	10-18
Expression Data Viewer	10-20
Exporting Switch Lists	10-23
Importing Significant Clusters	10-24
Assess Data Quality viewer	10-25

Getting started with Expression Analysis

The Expression Analysis tool enables you to perform the following tasks with ProteinLynx Global SERVER:

- Take mass spectrum data from samples labeled with different mass tags.
- Identify and extract pairs of labeled masses.
- Compute their relative abundance.
- Indicate whether they are upregulated or downregulated.

A wizard simplifies the complex of setting up an Expression analysis experiment. The wizard takes you through the process of specifying your samples and settings.

Note: The Expression software can be used as part of the optional Waters Protein Expression System. The Waters Protein Expression System provides a number of additional features, including label-free analysis. For more information, refer to the Waters Protein Expression System Operator's Guide.

To open the Expression Analysis tool, click the Expression Analysis icon .

Opening a project

Before creating an Expression analysis, you must create a project (see Creating a new project on page 3-2). To open a project that you have created, click the drop-down list in the toolbar, and then clicking the project you wish to open.

Experiment Analysis Design Manager

The Experiment Analysis Design Manager leads you through the creation of an Expression experiment.

To create a new Expression experiment:

1. Click Expression Analyses.
2. Right-click, and then click New Expression Analysis.

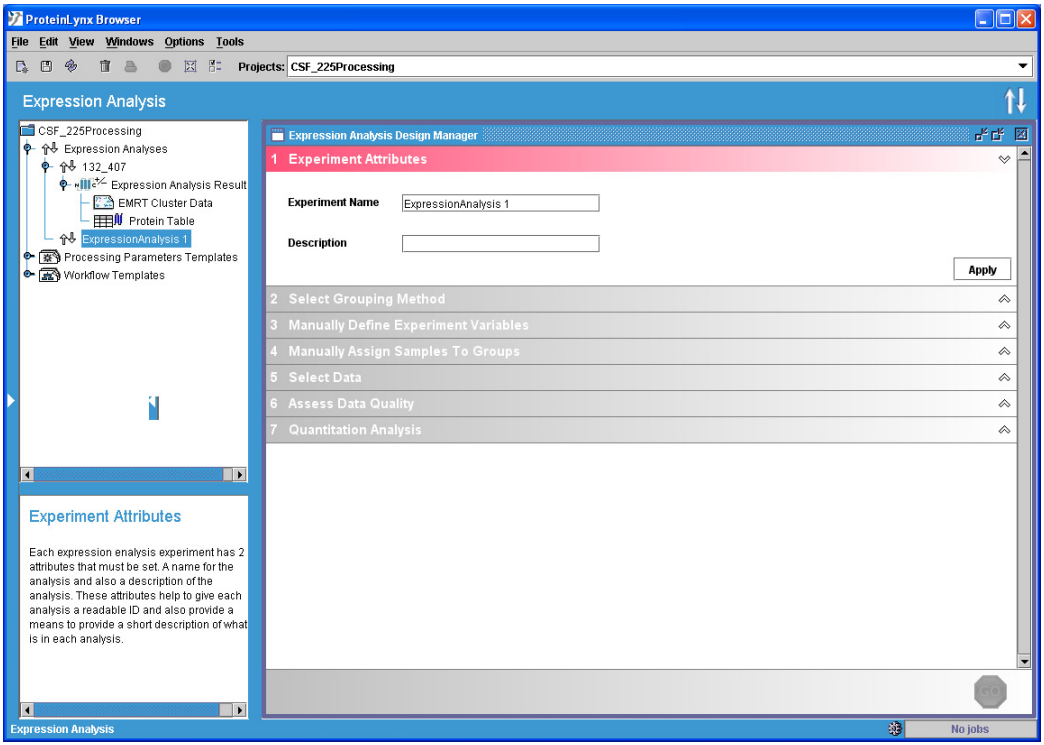
Result: A new Expression analysis is created in the tree, and the Design Manager opens at the first stage – Experiment Attributes.

To open an existing Expression experiment:

1. Click the name of the experiment.
2. Right-click, and then click Open Expression Analysis.

Result: The Design Manager opens at the section that needs your attention next.

Expression Analysis Design Manager:


ProteinLynx Browser
File Edit View Windows Options Tools
Projects: CSF_225Processing
Expression Analysis
CSF_225Processing
Expression Analyses
132_407
Expression Analysis Result
EMRT Cluster Data
Protein Table
ExpressionAnalysis 1
Processing Parameters Templates
Workflow Templates
Experiment Attributes
Each expression analysis experiment has 2 attributes that must be set. A name for the analysis and also a description of the analysis. These attributes help to give each analysis a readable ID and also provide a means to provide a short description of what is in each analysis.
Expression Analysis Design Manager
1 Experiment Attributes
Experiment Name
ExpressionAnalysis 1
Description
Apply
2 Select Grouping Method
3 Manually Define Experiment Variables
4 Manually Assign Samples To Groups
5 Select Data
6 Assess Data Quality
7 Quantitation Analysis
GO
Expression Analysis
No jobs

Note the following details, which apply to the Design Manager's seven sections.

- A red title indicates the section that needs completing next.
- A blue title indicates that the section is active, but that another section should be completed first.
- To apply the values that you specified for a section and progress to the next step, click Apply.
- To see or edit the values of another section, click the arrow at the right of the section heading. Click the arrow again to hide the section.

Experiment Attributes

This section names the Expression analysis, and specifies a description of its purpose.

Select Grouping Method

Use this section to specify how samples should be grouped. Groups are compared against one another.

Choose the processed sample (see Generating processed samples on page 4-5) that contains the samples you wish to use in the experiment.

If the optional Waters Protein Expression System is used, you can clear the 'Use isotope-labelled sample box' and choose any sample – not just processed samples.

Rules for isotope-labeled experiments:

- Only samples that have been labeled in Sample Manager using the Tag field appear in the drop-down list.
- Grouping methods other than placing the samples into separate groups are available only when there are more than two samples in the processed sample selected.

Grouping methods:

Method	How to	Result
Place samples into separate groups	1. Click 'Place samples in separate groups'. 2. In the list, click the samples you want to include in the analysis. Click [Select All] to include all the samples, or use Ctrl or Shift to select multiple samples. 3. Click Apply.	Each sample is in its own group.

Grouping methods:

Method	How to	Result
Group by experiment variable	1. Click 'Group by experiment variable'. 2. Click the sample variable (or attribute) by which you want to group. Custom attributes are included in this list. 3. To group by more than one attribute (so that samples which have the same values for Condition and Sex are grouped, for example) use Ctrl or Shift to select multiple attributes. 4. Click Apply.	Samples that share the selected variable are grouped together.
Manually assign sample groups	1. Click 'Manually assign sample groups'. 2. Click Apply, and then fill in the details in the Manually Define Experiment Variables section, described below.	Samples are grouped manually, according to user-defined variables.

Manually Define Experiment Variables

Use this section to define the variables you will use to group the samples.

Rules:

- This section applies only if 'Manually assign sample groups' is selected in the Select Grouping Method section.
- If manual group assignment is selected, at least one variable must be defined for each experiment.

To create a new variable:

1. Click New.
2. In the Variable box, type a name for the variable.
3. In the Values box, type a value for the variable.

4. Click Add.

To add values to a variable:

1. Click New.
2. In the Variable box, select the variable you wish to add a value for.
3. In the Values box, type a value for the variable.
4. Click Add.

Manually Assign Samples To Groups

Use this section to assign samples to groups, using the variables and values defined in Manually Define Experiment Variables.

Rule: This section applies only if 'Manually assign sample groups' is selected in the Select Grouping Method section.

To assign samples to groups:

1. Click the Variable drop-down box, and then click the variable you wish to group by.
2. Click the Value drop-down box, and then click the value appropriate for the samples you wish to assign.
3. In the Available Samples box, click the sample you wish to assign. More than one sample can be selected, using the Ctrl and Shift keys.
4. Click the >> button to add samples to the group.

Result: The selected samples are added to the Samples in Group box. They are made unavailable in the Available Samples box, and cannot be added to another group.

Select Data

The Select Data section shows the processed data associated with the samples identified in the previous sections. The first table contains a row for each sample; the second table contains a row for each replicate associated with the selected sample.

To show the attributes for a group:

1. In the group table, click the header of the third column.
2. In the drop-down list, click the attribute you wish to display.

To select data for inclusion in the experiment:

1. Click a group to see the associated replicates.
2. To include the replicate in the experiment, select the box in the Include column. To exclude the replicate, clear this box.

   Requirement: At least one replicate must be included for each group.
3. Repeat for other groups and replicates.

When Apply is clicked in this panel, the EMRTs (Exact Mass Retention Times) and Proteins are collated.

Results:

- A new results node appears below the node for the Expression analysis you are creating.
- For each replicate, an icon for the processed spectrum and an icon for the databank search are displayed. Click these icons to launch separate windows containing this information.

Assess Data Quality

This section usually becomes important only if you are unsure that the data is of good enough quality to use for quantitation. Clicking Apply in the Select Data section takes you directly to Quantitation Analysis.

The Assess Data Quality section contains a table with a row for each sample group. The table contains four columns – Group, Sample, Age, and Data View. The Data View column contains both a bar chart and a scatter chart icon.

Click either of these icons to display the Assess Data Quality viewer. See Assess Data Quality viewer on page 10-25 for further details.

Quantitation Analysis

Clicking Apply in the Select Data section brings you directly to this section.

Depending on the data selected, processing can take some time. Until processing is complete, some options in this section are unavailable. The progress of the processing can be monitored in the bottom right corner of the ProteinLynx browser.

Specify the type of data table you wish to generate from the analysis:

- EMRT (Exact Mass Retention Time) – processed data
- Proteins – results of searches

Depending on the other options selected for your experiment, this section can display options for specifying which normalization method to employ in the analysis – Automatic, Internal Standards, or no normalization.

If you wish to use Internal Standards, select the boxes beside the standards you want to use.

If you do not wish to use normalization at all, clear the Use Normalisation box.

The Go button is enabled when Apply is clicked in this section.

Starting an Expression analysis

Once the Expression analysis is configured in the Design Manager, the GO button

becomes available. Click GO to start the analysis.

Result: Once the analysis has completed, the tables specified in the Quantitation Analysis section are displayed. The quantitation can take some time – progress can be monitored in the bottom right corner of the ProteinLynx browser.

Viewing Expression Results

Expression results are automatically displayed when quantitation is completed.

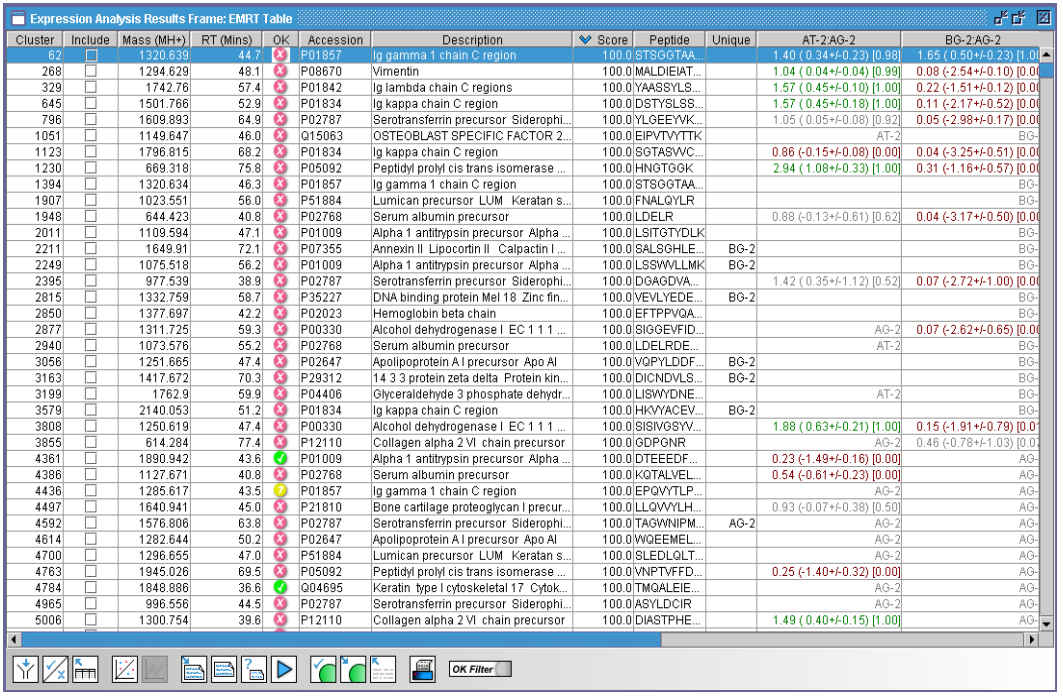
To display existing Expression results:

1. Expand the Expression experiment node.
2. Expand the Expression Analysis Result node.
3. Click EMRT Table or Protein Table, and then right-click.
4. Click Open Expression Table.

EMRT table

The table contains a number of columns, and a row for each cluster. Rows representing internal standards are shown highlighted in yellow.

EMRT table:



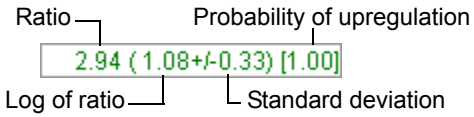
Cluster	Include	Mass (MH+)	RT (Mins)	OK	Accession	Description	Score	Peptide	Unique	AT-2:AG-2	BG-2:AG-2
62		1320.639	44.7		P01857	Ig gamma 1 chain C region	100.0	STSGGTAA...		1.40 (0.34+/-0.23) [0.98]	1.65 (0.50+/-0.23) [1.0
268		1294.629	48.1		P08670	Vimentin	100.0	MALDIEIAT...		1.04 (0.04+/-0.04) [0.99]	0.08 (-2.54+/-0.10) [0.00
329		1742.76	57.4		P01842	Ig lambda chain C regions	100.0	YAASSYLS...		1.57 (0.45+/-0.10) [1.00]	0.22 (-1.51+/-0.12) [0.00
645		1501.766	52.9		P01834	Ig kappa chain C region	100.0	DSTYSLSS...		1.57 (0.45+/-0.18) [1.00]	0.11 (-2.17+/-0.52) [0.00
796		1609.893	64.9		P02787	Serotransferrin precursor Siderophi...	100.0	YLGEEYVK...		1.05 (0.05+/-0.08) [0.92]	0.05 (-2.98+/-0.17) [0.00
1051		1149.647	46.0		Q15063	OSTEOBLAST SPECIFIC FACTOR 2...	100.0	EIPVTVYTTK		AT-2	BG-
1123		1796.815	68.2		P01834	Ig kappa chain C region	100.0	SGTASVVC...		0.86 (-0.15+/-0.08) [0.00]	0.04 (-3.25+/-0.51) [0.00
1230		669.318	75.8		P05092	Peptidyl prolyl cis trans isomerase ...	100.0	HNGTGGK		2.94 (1.08+/-0.33) [1.00]	0.31 (-1.16+/-0.57) [0.00
1394		1320.634	46.3		P01857	Ig gamma 1 chain C region	100.0	STSGGTAA...			BG-
1907		1023.551	56.0		P51884	Lumican precursor LUM Keratan s...	100.0	FNALQYLR			BG-
1948		644.423	40.8		P02768	Serum albumin precursor	100.0	LDELR		0.88 (-0.13+/-0.61) [0.62]	0.04 (-3.17+/-0.50) [0.00
2011		1109.594	47.1		P01009	Alpha 1 antitrypsin precursor Alpha ...	100.0	LSITGTYDLK			BG-
2211		1649.91	72.1		P07355	Annexin II Lipocortin II Calpactin I ...	100.0	SALSGHLE...	BG-2		BG-
2249		1075.518	56.2		P01009	Alpha 1 antitrypsin precursor Alpha ...	100.0	LSSWVLLMK	BG-2		BG-
2395		977.539	38.9		P02787	Serotransferrin precursor Siderophi...	100.0	DGAGDVA...		1.42 (0.35+/-1.12) [0.52]	0.07 (-2.72+/-1.00) [0.00
2815		1332.759	58.7		P35227	DNA binding protein Mel 18 Zinc fin...	100.0	VEVLYEDE...	BG-2		BG-
2850		1377.697	42.2		P02023	Hemoglobin beta chain	100.0	EFTPPVQA...			BG-
2877		1311.725	59.3		P00330	Alcohol dehydrogenase I EC 1 1 1 ...	100.0	SIGGEVFID...		AG-2	0.07 (-2.62+/-0.65) [0.00
2940		1073.576	55.2		P02768	Serum albumin precursor	100.0	LDELRDE...		AT-2	BG-
3056		1251.665	47.4		P02647	Apolipoprotein A I precursor Apo AI	100.0	VQPYLDDF...	BG-2		BG-
3163		1417.672	70.3		P29312	14 3 3 protein zeta delta Protein kin...	100.0	DICNDVLS...	BG-2		BG-
3199		1762.9	59.9		P04406	Glyceraldehyde 3 phosphate dehydr...	100.0	LISWYDNE...		AT-2	BG-
3579		2140.053	51.2		P01834	Ig kappa chain C region	100.0	HKVYACEV...	BG-2		BG-
3808		1250.619	47.4		P00330	Alcohol dehydrogenase I EC 1 1 1 ...	100.0	SISIVGSYV...		1.88 (0.63+/-0.21) [1.00]	0.15 (-1.91+/-0.79) [0.0
3855		614.284	77.4		P12110	Collagen alpha 2 VI chain precursor	100.0	GDPGNR		AG-2	0.46 (-0.78+/-1.03) [0.0
4361		1890.942	43.6		P01009	Alpha 1 antitrypsin precursor Alpha ...	100.0	DTEEEDF...		0.23 (-1.49+/-0.16) [0.00]	AG-
4386		1127.671	40.8		P02768	Serum albumin precursor	100.0	KQTALVEL...		0.54 (-0.61+/-0.23) [0.00]	AG-
4436		1285.617	43.5		P01857	Ig gamma 1 chain C region	100.0	EPQVYTLP...		AG-2	AG-
4497		1640.941	45.0		P21810	Bone cartilage proteoglycan I precur...	100.0	LLQVVYLH...		0.93 (-0.07+/-0.38) [0.50]	AG-
4592		1576.806	63.8		P02787	Serotransferrin precursor Siderophi...	100.0	TAGWNIPM...	AG-2	AG-2	AG-
4614		1282.644	50.2		P02647	Apolipoprotein A I precursor Apo AI	100.0	WQEEMEL...		AG-2	AG-
4700		1296.655	47.0		P51884	Lumican precursor LUM Keratan s...	100.0	SLEDLQLT...		AG-2	AG-
4763		1945.026	69.5		P05092	Peptidyl prolyl cis trans isomerase ...	100.0	VNPTVFFD...		0.25 (-1.40+/-0.32) [0.00]	AG-
4784		1848.886	36.6		Q04695	Keratin type I cytoskeletal 17 Cytok...	100.0	TMQALEIE...		AG-2	AG-
4965		996.556	44.5		P02787	Serotransferrin precursor Siderophi...	100.0	ASYLDCIR		AG-2	AG-
5006		1300.754	39.6		P12110	Collagen alpha 2 VI chain precursor	100.0	DIASTPHE...		1.49 (0.40+/-0.15) [1.00]	AG-

Sort the results by clicking a column heading. Click the heading again to reverse the order of the sort. To re-order the columns, click the heading and drag the column to the desired location.

For each comparison there is a column. The cells in these columns, when filled completely, contain this information:

- Ratio of Condition A:Condition B (a condition is sample or group of samples)
- Log of that ratio
- Standard deviation of the log
- Probability of upregulation

Typical comparison column cell:



The text is green if the probability of upregulation is 0.95 or more, and red if the probability is 0.05 or less. A value of 1.00 indicates that the cluster is definitely upregulated; a value of 0.00 indicates that the cluster is definitely downregulated.

If the cluster or protein only appeared in one of the conditions (groups) then the name of the group that it appeared in is displayed in the cell. If the item appeared in neither of the conditions, the cell is blank.

If the cluster or protein only appeared in one of the conditions (groups), and appeared in every injection for that group, the group's name is displayed in the Unique column.

To curate (organize) your data:

Rule: In the EMRT table, curation is possible only on clusters with identification information.

The following steps apply to curation in the EMRT table. In the Protein table, only step 3 applies.

1. Click the cluster of interest.
2. Click the Curate Data button, .

Result: The individual peptide identifications for the selected cluster are displayed in the upper half of Data Curation window. The lower half displays the high energy fragmentation data associated with the selected cluster.

3. To mark a protein or peptide, click the icon to cycle through OK, unsure, and not OK states.

4. When you are satisfied with your settings, click to close the window.

You can choose to show all clusters, those clusters marked as OK or unsure, or only those clusters marked as OK. To control which clusters are displayed, click

to cycle through the display modes.

To view the workflow for a cluster:

1. In the results table, click the cluster.

2. Click the Show Workflows button, .

Result: The workflow is displayed in the Results browser (see Viewing results in the Results Browser on page 6-1 for more information).

To view the replicates for a cluster or protein:

1. Click the line in the EMRT or Protein table representing the cluster or protein you wish to view the replicates for.

2. Click the Open Replicate Viewer button, .

Result: The replicates or peptides for the selected cluster or protein are displayed.

To export your data:

Tip: If there are many results, you might wish to filter the results (see page 10-13) before exporting them.

1. Click the Export Data button.

2. In the Export Data dialog box, select the boxes beside the columns you want to export, and clear the boxes beside the columns you do not want to export.

3. Click OK.
4. Type a name for the export file, and then click Save.

Result: A tab-delimited file is created with the specified name. If there are many results it can take a few moments for the export file to be created.

To print your data:

Tip: If there are many results, you might wish to filter the results (see page 10-13) before printing them.

1. Click the Print Data button. The Print Wizard (see Using print wizards on page 11-3) is displayed.
2. Follow the on-screen instructions in the Print Wizard, clicking Next to progress from one step to the next, and Finish to print.

To include/exclude all clusters:

To include all clusters, click .

To exclude all clusters, click .

Rule: Only one of these buttons is displayed at any one time. If the Include All button is clicked, the Exclude All button is then displayed. If the Exclude All button is clicked, the Include All button is then displayed.

Protein table

The Protein table is similar to the EMRT table (see page 10-10), but does not contain columns for Cluster, Include, Average Mass, Average RT, Peptide, or Probability.

Filtering the results

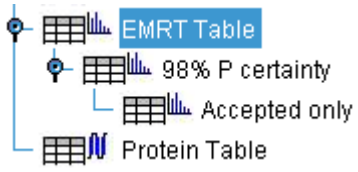
To make the results easier to interpret – or to reduce the size of the list in preparation for printing – you can generate new results tables, filtered by various criteria.

To filter the results:

1. Click the Filter button, .
2. Type a title for the results table that will be generated for the filtered results.
3. Set the filtering options as required (see Replicate filter on page 10-14, Confidence Limit, P value, and Ratio filters on page 10-15, and Additional Filter settings on page 10-15).
4. To see the data that will be included in the filtered results in the Log Plot Viewer (see page 10-18) click Preview. To generate the filtered results table, click OK.

Result: A new table is generated containing the filtered results. A node will be added to the navigation tree below the results table that has been filtered.

Example filtered results tree:



Rule: EMRT and Protein tables, including tables containing filtered results, cannot be deleted.

Replicate filter

The Replicate filter enables you to limit the results to a specified number of replicates per sample.

To set a replicate filter:

1. Select Use Replicate Filter Settings.
2. For each sample, set the maximum number of replicates that you want to be included for that sample in the filtered results. You can either type the limit directly in the Number of Replicates column or use the up and down arrows to increase or decrease the limit.

   Tip: To specify the same number of replicates for each sample, click the number in the 'Set the Number of Replicates in all' drop-down list.

Confidence Limit, P value, and Ratio filters

These filters enable you to return only those results that fall within set limits for the standard deviation of the log ratio, probability of upregulation, or ratio.

To set a confidence limit (standard deviation of the log ratio) filter:

1. Select Use Confidence Limit Settings.
2. Type a limit in the Ceiling box, or drag the slider to set a limit.

To set a probability of upregulation (P value) filter:

1. Select Use P > 1 Settings.
2. Type values in the boxes, or drag the sliders to set the limits. Clusters with P values between the Floor and Lower and clusters with P values between the Upper and Ceiling are included in the filtered results.

To set a ratio filter:

1. Select Use Ratio Settings.
2. Type values in the boxes, or drag the sliders to set the limits. Clusters with log ratios between the Floor and Lower and clusters with log ratios between the Upper and Ceiling are included in the filtered results.

Additional Filter settings

There are a number of additional ways of filtering your data. To enable these filters, click Use Additional Filter Settings.

Additional filters:

Filter	Effect
Display all items with the following OK level(s)	Only those clusters or proteins marked with the selected status (see To curate (organize) your data: on page 10-11) are included.
Remove all proteins with a score less than	Only proteins with a score higher than the value entered are included.

Additional filters:

Filter	Effect
Remove all EMRTs with an average mass error (PPM) less than	Only EMRTs with an average mass error (the root mean square, calculated in parts per million) greater than the value entered are included. Average mass errors are typically very small.
Remove all EMRTs with a percentage CV in retention time greater than	Only EMRTs with a coefficient of variation in retention time that is smaller than the value entered are included.
Remove all EMRTs with a percentage CV in intensity greater than	Only EMRTs with a coefficient of variation in intensity that is smaller than the value entered are included.

Importing workflows

Import workflows to apply the protein identification results of one or more databank searches to your EMRT results table.

To import workflows:

1. Click the Import Workflows button, .
2. In the Select Workflows dialog box, select the boxes on the rows relating to the workflows you wish to import.
3. Click OK.

Result: The protein IDs from the selected workflow(s) are imported into the EMRT result table, where appropriate. Importing can take some time – progress can be monitored in the bottom right corner of the ProteinLynx browser.

Searching EMRTs from the EMRT table

To search EMRTs:

1. In the EMRT results table, select the Include check box for each cluster you wish to search (to select all the clusters, see To include/exclude all clusters: on page 10-13).
2. Click the Set Databank Search Parameters button, .
3. Set the parameters as required (see Databank search parameters on page 14-5 for information on the options available).
4. Click to close the Databank Search parameters window.
5. Click the Submit Databank Search button, .
6. Type a title for the workflow, and then click OK.

Result: When the search is complete the protein identifications returned are automatically added to the EMRT table for the selected clusters. Searching can take some time – progress can be monitored in the bottom right corner of the ProteinLynx browser.

Log Plot Viewer

To open the Log Plot viewer, click .

To set the values for axes:

1. To set the values displayed on the *y* axis, click . To set the values for the *x* axis, click .
2. Click the values that you want to display on that axis.

To alter the range displayed on an axis:

1. To modify the lower limit of the range, click and hold the left or bottom axis slider. To modify the upper limit of the range, click and hold the right or top axis slider.

2. Drag the slider to modify the range limit.

To select data points:

1. Click one edge of the area you want to select.
2. To select a rectangular area, drag to the opposite corner of the area you want to select. To select an area freehand, hold down Shift while you draw the area you want to select.
3. When the correct area is highlighted, release the mouse button.

Result: The selected data points are shown in red. Click anywhere to deselect the points and start again.

To perform a databank search on selected data points:

1. Click the Set Databank Search Parameters button, .
2. Set the parameters as required (see Databank search parameters on page 14-5 for information on the options available).

Tip: It is advisable to specify a databank that contains the majority of protein sequences that could be in the sample data searched.

3. Click to close the Databank Search parameters window.
4. Click the Search selected items button, .
5. Type a title for the workflow, and then click OK.

Result: Protein identifications are returned for the selected EMRTs. Searching can take some time – progress can be monitored in the bottom right corner of the ProteinLynx browser.

To display only unique EMRTs:

Click the Unique EMRTs Only button, . To revert to displaying all the non-unique EMRTs, click .

To display each identified protein on a separate plot:

Click the Trellis data by protein id button, . Each identified protein is displayed in its own plot, and all unidentified proteins are displayed on one plot.

To copy the log plot to the clipboard:

Click the Copy button, . The log plot is copied to the Windows clipboard, from where it can be pasted into other applications.

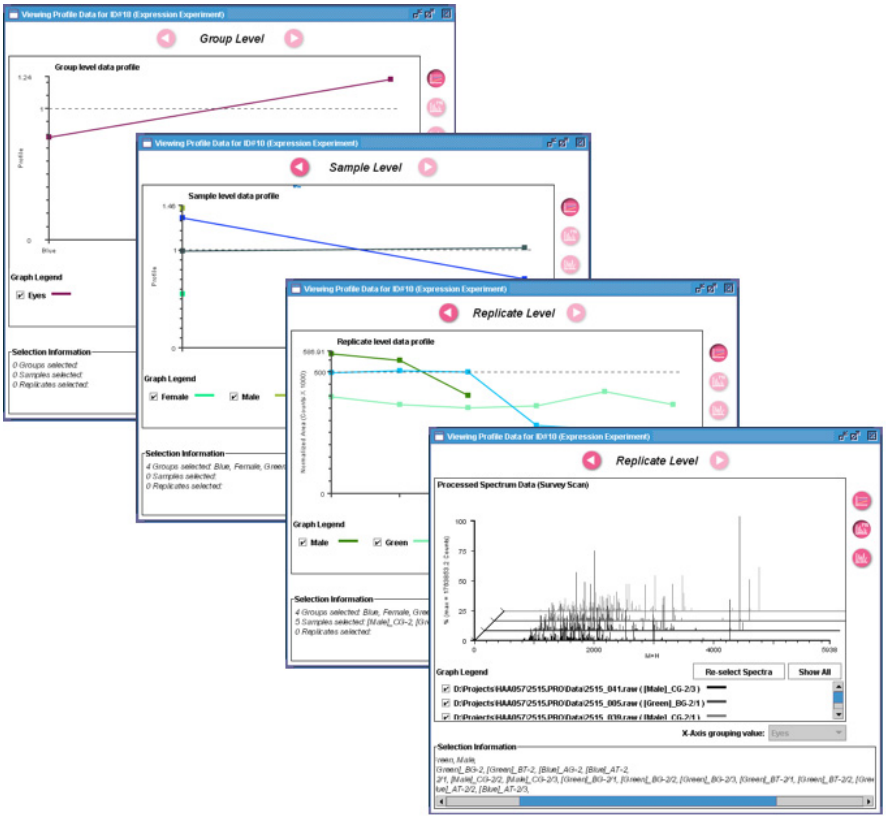
Expression Data Viewer

Use the Expression Data Viewer to view graphical representations of the relationships between groups, samples, and replicates. You can also view the raw and processed spectra associated with selected replicates.

To open the Data Viewer, click a row in the EMRT or Protein Table, and then click .

Rule: This button is not available if a unique protein is selected in the Protein Table.

Expression Data Viewer:



There are three levels of view available - Group level, Sample level, and Replicate/Spectrum level. At each level, a number of actions are possible:

- Control which groups, samples, or replicates are displayed by selecting or clearing the check boxes below the graph.
- Alter the *x*-axis value by clicking the X-Axis grouping value list, and then clicking the value you want to use.
- Select traces or points on the graph by dragging a rectangle over the points you want to select.

Group level

When the Data Viewer is opened, it usually appears at Group level. If one or more groups are selected, the icon is available. Click the icon to go to the Sample level for the selected groups.

Sample level

Click to go back to the Group level.

If one or more samples are selected, click to go to the Replicate/Spectrum level for the selected samples.

Replicate/Spectrum level

Rule: For isotopic (ICAT™,for example) and isobaric (iTRAQ™, for example) experiments this level is labeled Spectrum level. For other experiment types, it is labeled Replicate level. In either case. the operations available remain the same.

Click to go back to the Sample level.

If one or more replicates are selected, the Show Processed Data and Show Raw Data icons become available.

To display raw or processed data:

1. In the Replicate level graph, select traces or points by dragging a rectangle over the points you want to select.

2. Click to display processed data, or to display raw data.
3. Select the check boxes beside the replicates you wish to view spectra for.
4. Click Show Selected.

Result: The selected spectra are displayed on a single graph.

To show or hide the spectra on the graphical display, select or clear the check boxes in the Graph Legend section.

To select different replicates for display, click Re-select Spectra, and then repeat steps 3 and 4 above.

To switch back to the data profile view, click .

Tip: Switching to the profile view does not reset your spectra selections. Click the appropriate icon to revert to the spectra view.

Exporting Switch Lists

Clusters can be exported from EMRT results tables as switch lists.

To export clusters as a switch list:

1. In the EMRT table (see EMRT table on page 10-10) select the check box in the Include column beside each cluster you wish to include. See To include/exclude all clusters: on page 10-13 for details on including all clusters.
2. Click Export Switch List, .
3. In the Export Switch List dialog box, browse to the location you wish to save the file in, and then type a name for the switch list file.
4. Click Save.

Result: A text file, containing the switch list information for the selected clusters, is created in the location specified.

Importing Significant Clusters

You can import a list of significant clusters into your EMRT results table to simplify and accelerate the process of selecting clusters for other operations, such as exporting switch lists or searching EMRTs.

To import significant clusters:

1. In the EMRT results table (see EMRT table on page 10-10), click Import Significant Clusters, .
2. Browse to the location of the clusters file you wish to import.
3. Click the file, and then click Open.

Result: The Include column is selected for the clusters listed in the imported file.

Significant clusters list file format

Significant cluster list files are plain text files, containing one cluster number on each line.

Example:

```
4
18
41
55
84
101
142
165
```

Assess Data Quality viewer

If you are unsure whether your data is good enough for quantitation – or if you find that your quantitation results are not what you expect – you can view statistics for each injection in the Assess Data Quality viewer.

To open the Assess Data Quality viewer:

1. Click the arrow at the right side of the Assess Data Quality section so that the panel is displayed.

   Requirement: You must have an Expression experiment open to do this. See Experiment Analysis Design Manager on page 10-3 for details.

2. In the Data View column, click either the bar chart or scatter chart icon.

To set the values for axes:

1. To set the values displayed on the *y* axis, click . To set the values for the x axis, click .
2. Click the values that you want to display on that axis.

To alter the range displayed on an axis:

1. To modify the lower limit of the range, click and hold the left or bottom axis slider. To modify the upper limit of the range, click and hold the right or top axis slider.

2. Drag the slider to modify the range limit.

To switch between bar chart and scatter chart view:

Click to show the bar chart view.

Click to show the scatter chart view.

To show/hide the EMRT and Peptide panes:

Click to show/hide the EMRT Clusters pane.

Click to show/hide the Matching Peptides pane.

11 Creating print templates and printing project data

The Print Tool enables the creation and modification of printing templates. Printing templates are used to control how project data is printed.

Contents:

Topic	Page
Printing data	11-2
Using print wizards	11-3
Opening and deleting print templates	11-12
Creating print templates	11-13
Customizing print templates	11-19

Printing data

When you print data you combine project or workflow data with a template. Rendering combines the template and the data to produce a printed report, a preview, or an exported file. Files can be exported as two types:

- Comma-separated values files (*.csv)
- HTML files (*.html).

There are default templates supplied with PLGS, or you can create your own using the Print Tool. The Print Tool enables you to create, modify, and preview two types of template:

- Project template – Prints details of all the hits in the project that have a score of higher than zero.
- Workflow template – Prints all details of the workflow used to obtain the data, and a sorted list of proteins, peptides, and possibly masses.

Recommendation: New users should use the default templates supplied. If you are creating a template, open a default template and save it as a new template. Then edit the text, graphics, and so on.

The template editor enables you to edit and create print templates using an WYSIWYG (What You See Is What You Get) interface. You use a properties editor to edit objects: paragraphs, images, and so on. Results pages are organized into hierarchical trees, where you can apply limiting and sorting, and then preview with the standard results set or any of your project data.

There are print wizards to print the data. The print wizards are accessed from the navigator trees, toolbar, or results windows within the PLGS tools. You can print project data from the navigator tree of any tool that shows the project name. However, you can only print workflow results from the Container Manager navigator tree or a results window.

Note: The speed of rendering depends on the amount of data being applied to the template and the specification of the computer.

Using print wizards

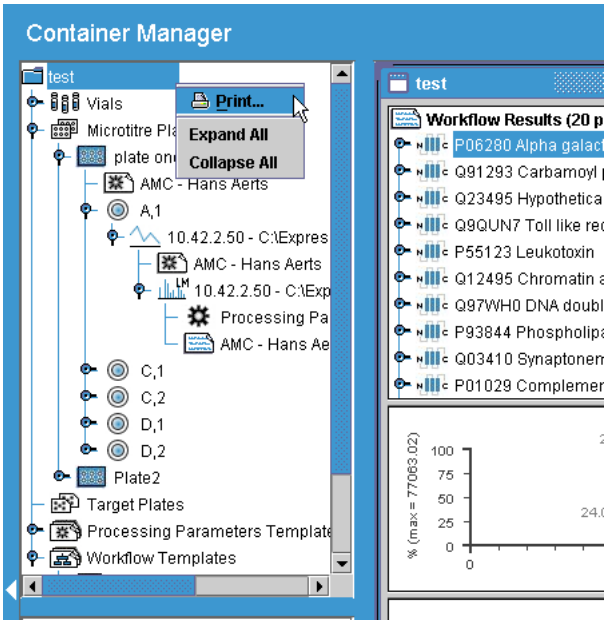
To print project or workflow data, you use the project or workflow print wizards. You can print project data from the navigator tree of any tool that shows the project name. However, you can only print workflow results from the Container Manager navigator tree or a results window.

Project print wizard

To use the project print wizard:

1. In the navigator tree of any tool, click the project name, and then right-click.

   Project print wizard - pop-up menu in navigator tree:

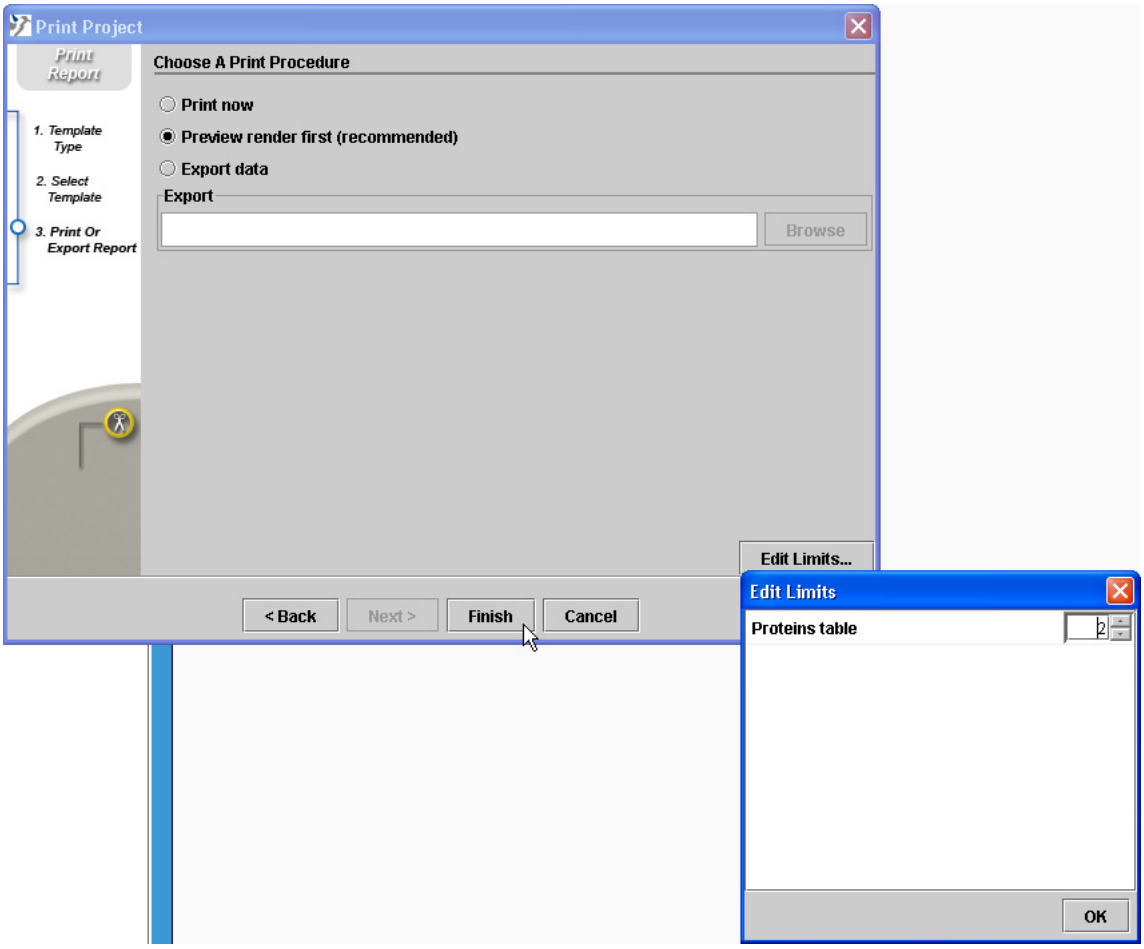


2. Click Print.
3. Select either default templates or user-defined templates, and then click Next to open a template selection dialog box.

   Recommendation: New users should use default templates.

4. Click a suitable template, and then click Next.

Project print wizard - Choose a Print Procedure:



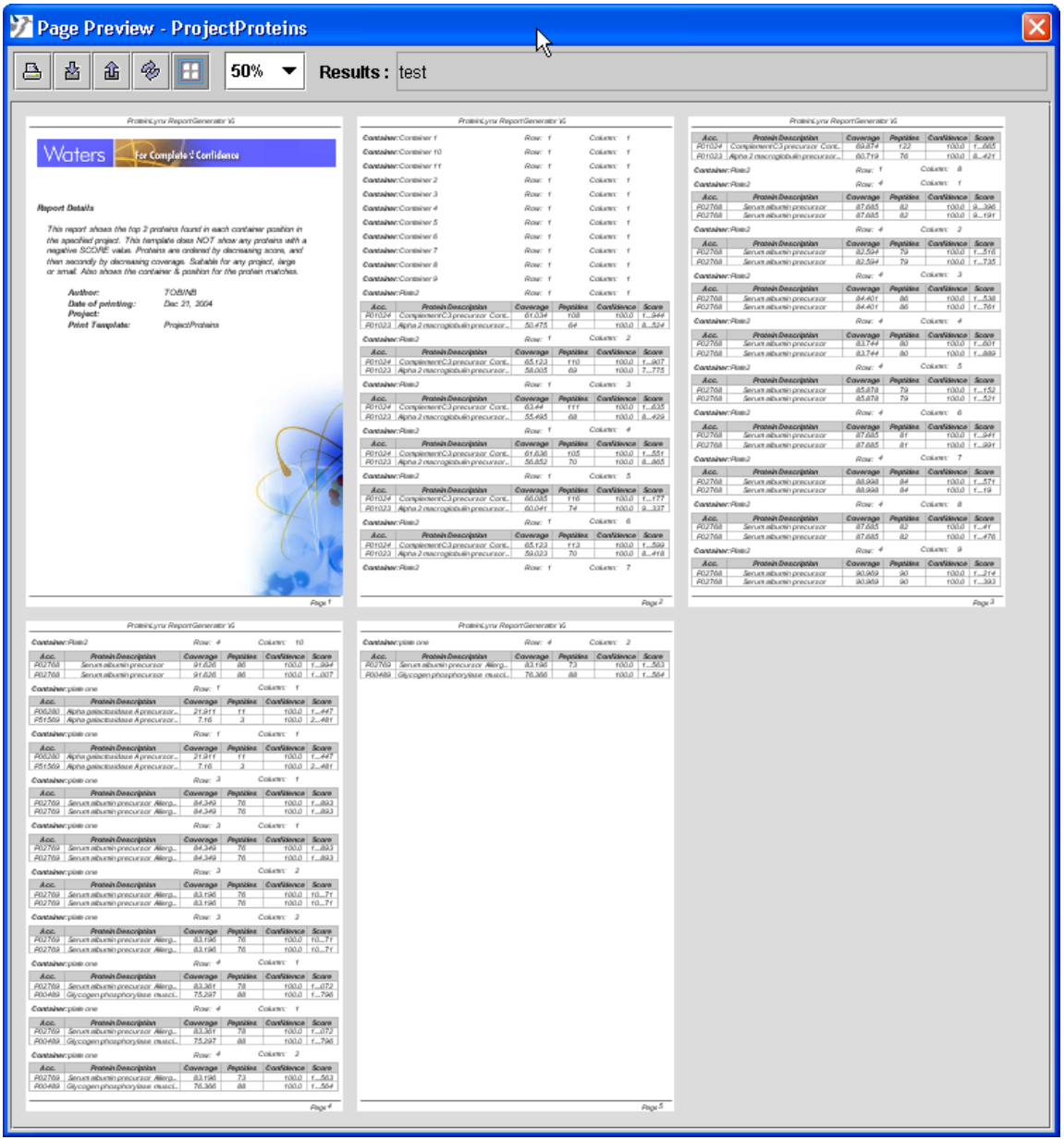
In this screen, you can print immediately, preview the report (see Figure titled “Previewing a project report:” on page 11-5) or export the data to a *.csv or *.html file type.

Recommendation: It is recommended that you preview the report.

The Edit Limits dialog box enables you to override the limiting options for the results that are set in the template (see Limiting results on page 11-17). However, the settings in this dialog box are not saved in the template.

5. After selecting an option, click Finish.

Previewing a project report:


Page Preview - ProjectProteins
50%
Results :
test
Waters
For Complete Confidence
Report Details
This report shows the top 2 proteins found in each container position in the specified project. This template does NOT show any proteins with a negative SCORE value. Proteins are ordered by decreasing score, and then secondly by decreasing coverage. Suitable for any project, large or small. Also shows the container & position for the protein matches.
Author: TOBIAB
Date of printing: Dec 21, 2004
Project:
Print Template: ProjectProteins

The toolbar has various functions.

Print preview toolbar functions:

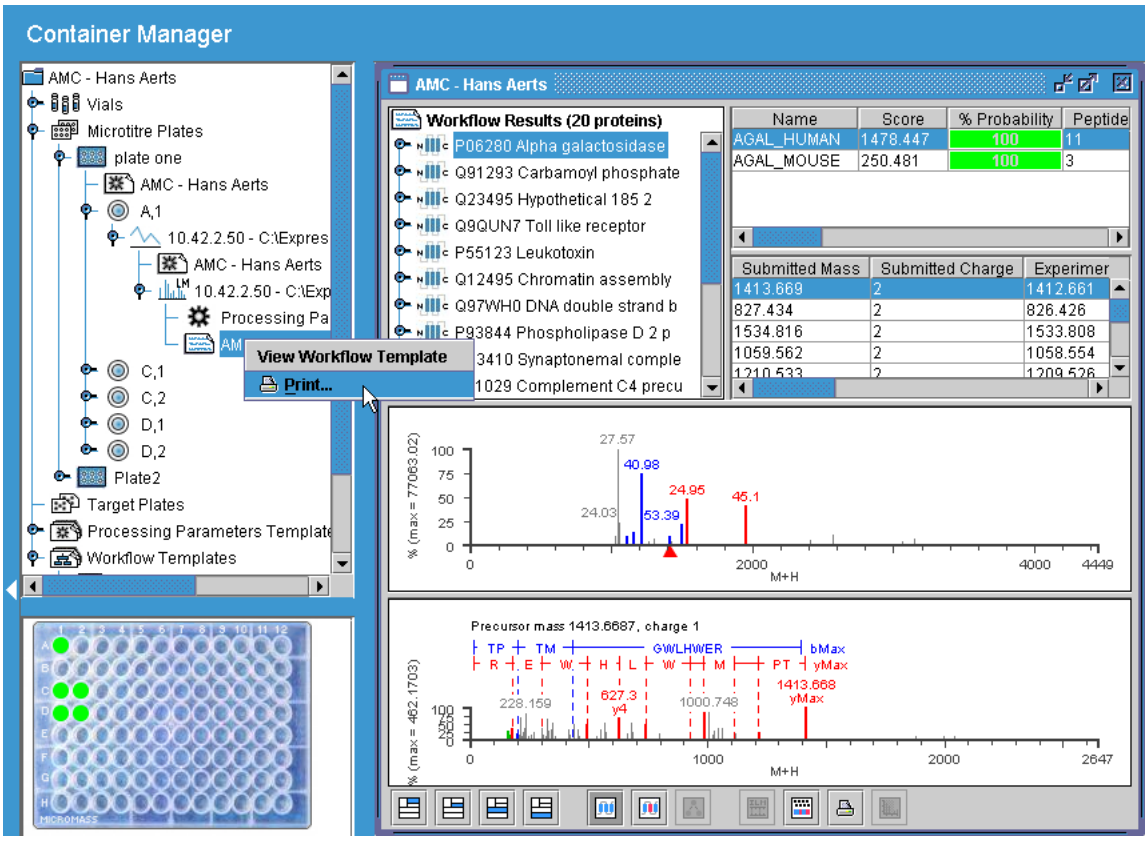
Function	Description
Print	Print the project from this screen.
Import	Import another project to be previewed, printed or exported.
Export	Export the project results to a *.csv or *.html file.
Refresh	Refresh the preview.
Toggle grid	Preview pages horizontally across the display.
Zoom	Increase or decrease the scale of the view (range = 25% to 200%). Use this with the Toggle grid function to display pages across the display, as in the graphic.

Workflow print wizard

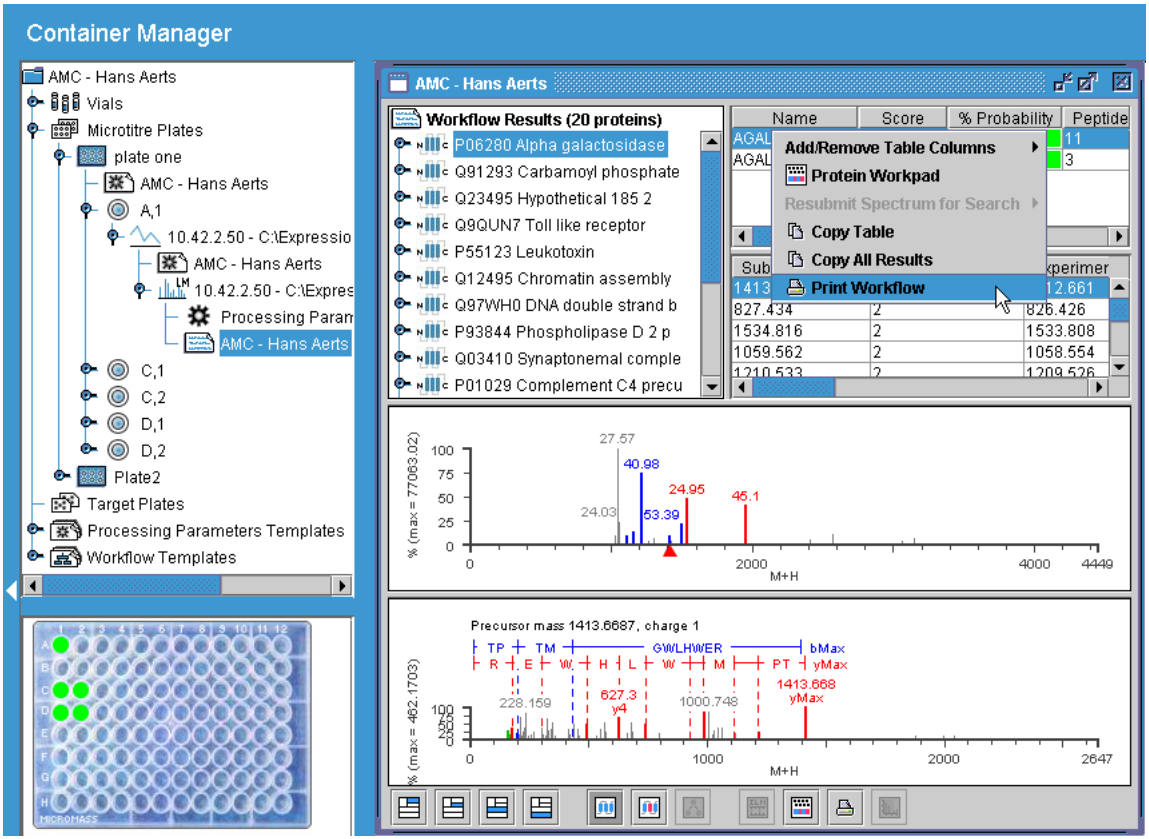
To use the workflow print wizard:

1. You can open a workflow print wizard in two ways:
   - In the Container Manager navigator tree, click the workflow results (not the workflow template), and then right-click. Click Print.
   - In the results table, click a protein, and then right-click. Click Print Workflow.

Workflow print wizard - pop-up menu in Container Manager navigator tree:



Workflow print wizard - pop-up menu in a results table:



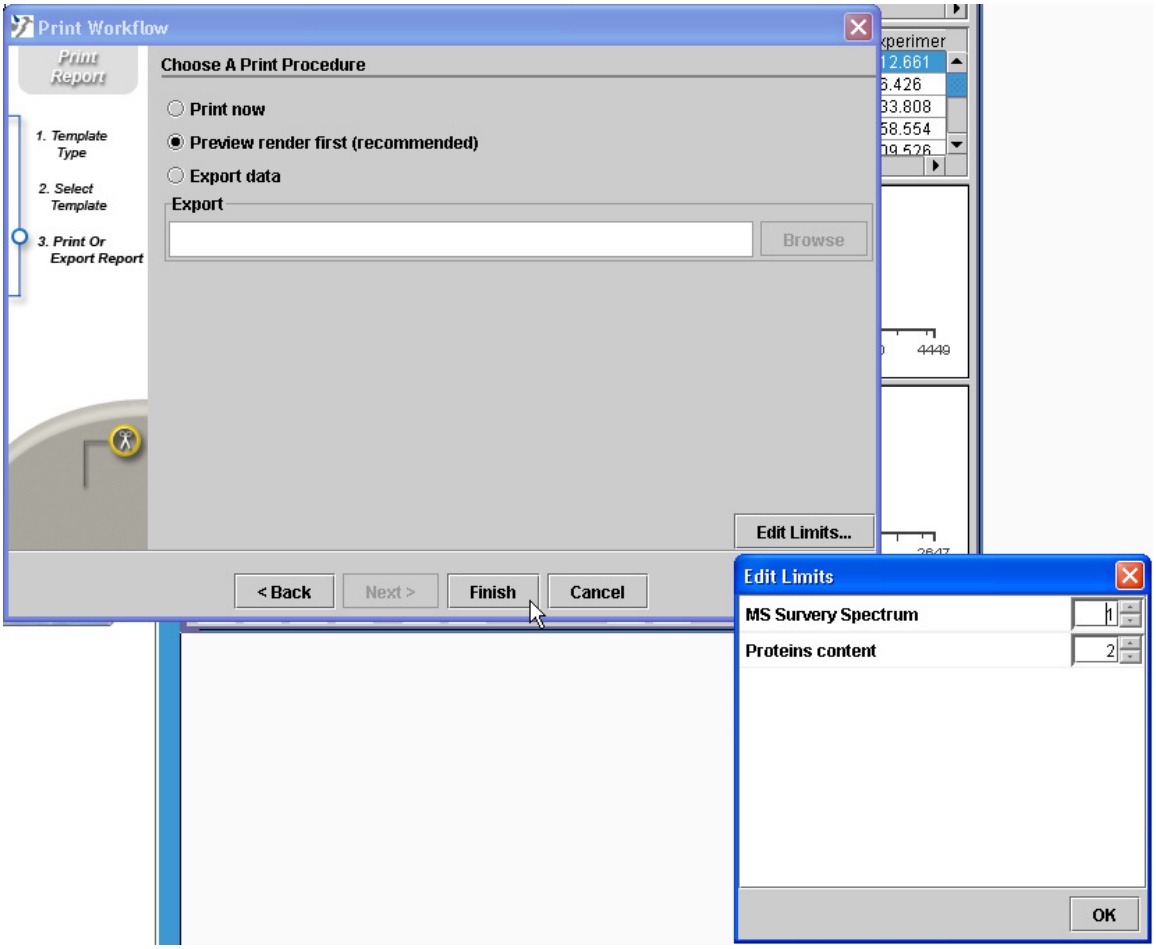
Whichever method is used, a template selection dialog box opens.

2. Select to use either default templates or user-defined templates, and then click Next.

   Recommendation: New users should select default templates.

3. Click a suitable template, and then click Next.

Workflow print wizard - Choose a Print Procedure:



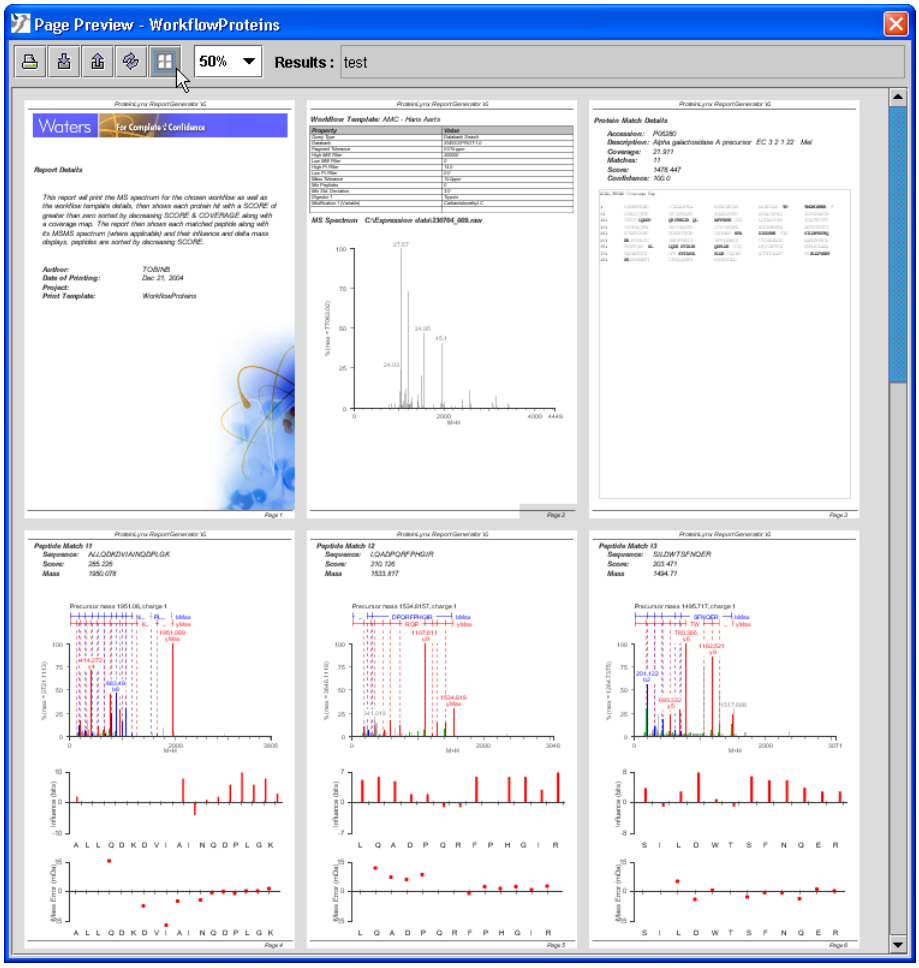
In this screen, you can print immediately, preview the report (see Figure titled “Previewing a Workflow report:” on page 11-10) or export the data to a *.csv or *.html file type.

Recommendation: It is recommended that you preview the report.

The Edit Limits dialog box enables you to override the limiting options for the results that are set in the template (see Limiting results on page 11-17). However, the settings in this dialog box are not saved in the template.

4. After selecting an option, click Finish.

Previewing a Workflow report:



The toolbar has various functions.

Print preview toolbar functions:

Function	Description
Print	Print the workflow from this screen.
Import	Import another workflow to be previewed, printed or exported.
Export	Export the workflow results to a *.csv or *.html file.

Print preview toolbar functions: (Continued)

Function	Description
Refresh	Refresh the preview.
Toggle grid	Preview pages horizontally across the display.
Zoom	Increase or decrease the scale of the view (range = 25% to 200%). Use this with the Toggle grid function to display pages across the display, as in the graphic.

Opening and deleting print templates

The same dialog box is used to open or delete existing templates, whether they are default or user-defined.

To open or delete an existing template:

1. In the tool tray, click to open the Print Tool.
2. Click .

   Alternative: Click File > Open.
3. Click the template name.
4. To open the template, click Open. To delete the template, click .

Creating print templates

Use the Print Tool to create project or workflow templates. The templates you produce are displayed as user-defined templates in the Project print wizard or Workflow print wizard.

See also:

- Project print wizard on page 11-3.
- Workflow print wizard on page 11-6

To open the Print Tool, click the Print Tool icon in the tool tray.

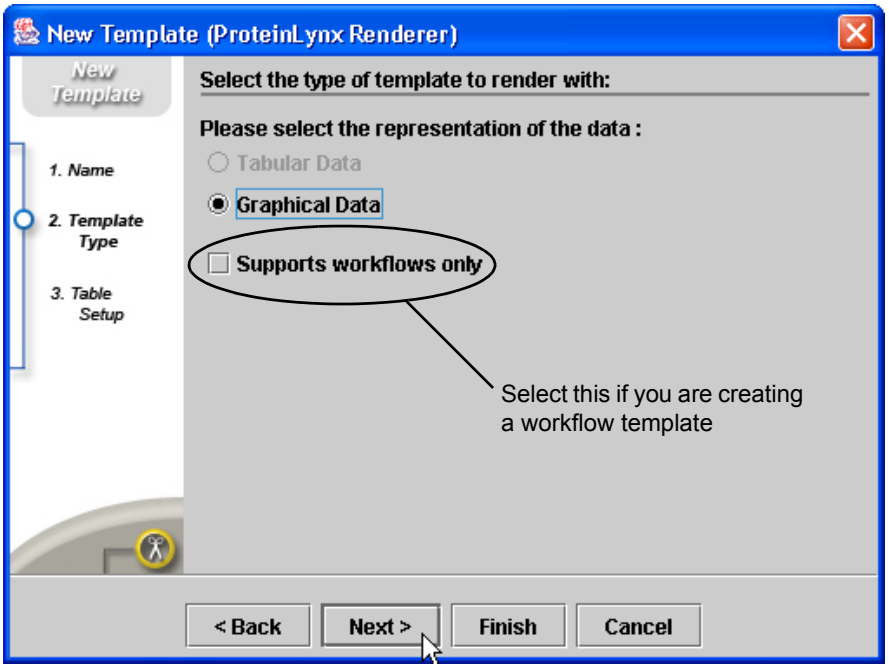
To create a new template:

1. Click .

   Alternative: Click File > New.

2. Type a name, and then click Next.

Print Tool - New Template:



3. Select either Graphical Data or Tabular Data.
4. Choose a setting for the Support workflows only check box:
   - For a template to print data for a whole project, clear the box.
   - For a template to print data for specific workflows only, select the box.

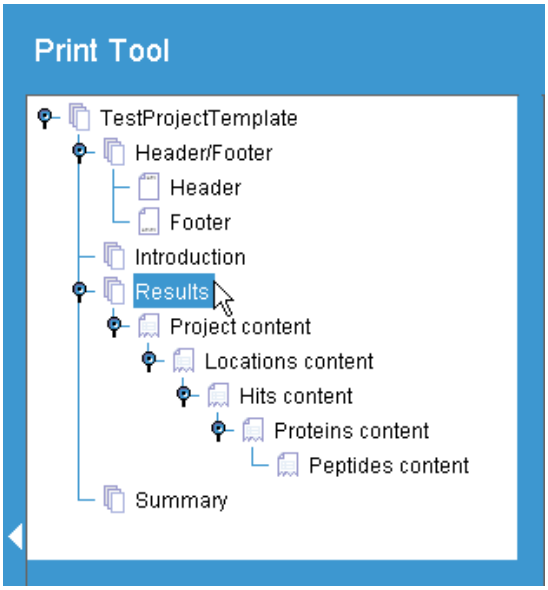
   Rule: The Tabular Data option is only available if you have set up the printing preferences to enable quick table pages. See Printing tab on page 2-16 for details.
5. Click Next, and then select the ways that you want information to be grouped.

   Tip: The selections that you make are displayed in the Results section of the template navigator tree in the same order as they are displayed in these screens.
6. Click Next, and then select the data sets to be displayed. You can change the order of the data sets by using the up and down arrows.
7. Click Finish.

Results:

- The template details are shown in the Print Tool view in the browser.
- The Table Setup selections are chained in the Results section of the template navigator tree.

Print Tool - Table Setup - display in the navigator tree of a template:



You can still add content to the Results section after the template has been created.

8. Click to save the template.

Adding content to the results nodes

You can add content for the grouping and data sets from within the template navigator tree.

To add content from the template navigator tree:

1. In the template navigator tree, right-click the Results node or one of the content nodes.
2. Click Insert > Content Page.
3. Select a results table, and then click OK.

Result: The content page is added to the navigator tree below the selected node.

Filtering, sorting and limiting in results nodes

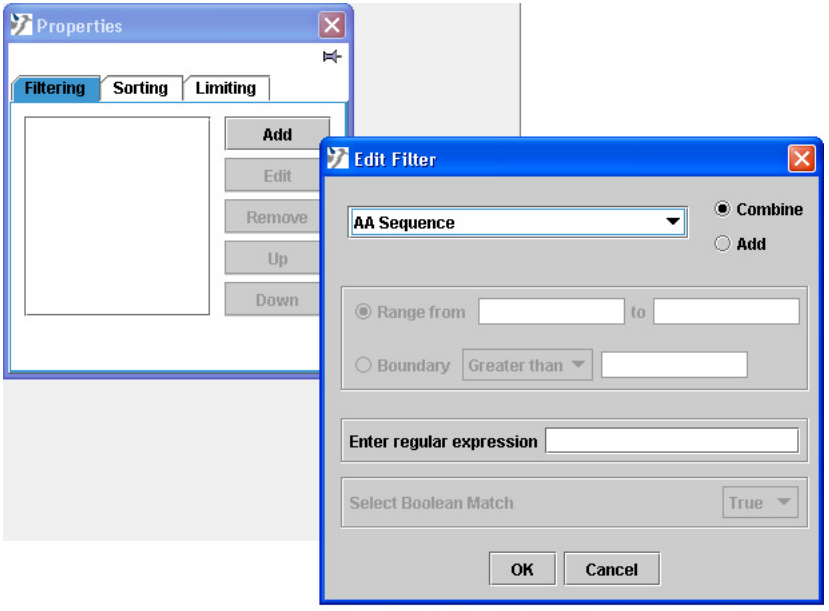
You can filter, sort, and limit the results in content pages.

In the navigator tree, click a content page. Properties for the content page are displayed in the lower part of the pane.

Filtering results

To add filters, click the Filtering tab, and then click Add.

Properties dialog box - adding a filter:



The drop-down menu contains all the fields available to filter the results.

Different options are available in the dialog box, depending on the field selected:

- Numeric – Range and Boundary options are enabled.
- Text – Enter regular expression option is enabled.
- Curated – Select Boolean Match is enabled.

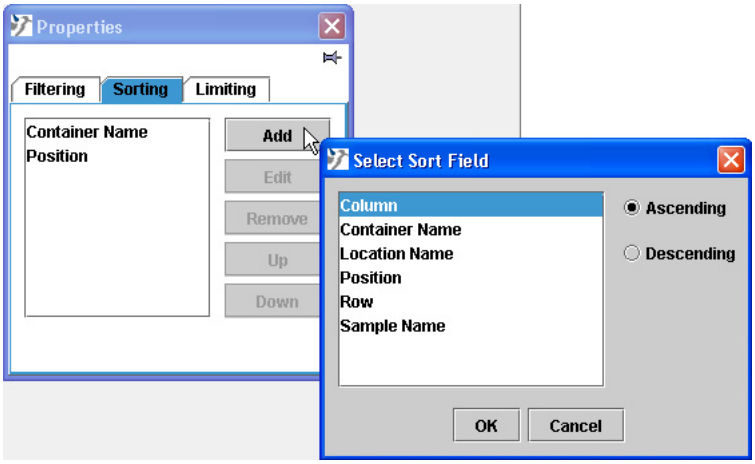
The Combine and Add options enable you to either combine this filter with other filters, or use this filter in addition to other specified filters.

Example: If you apply two combined filters to the results, the report only shows a condition (for example, a protein) that satisfies both filters; if the same two filters are applied as additional, the protein is shown if it satisfies either filter.

Sorting results

To add sorting fields, click the Sorting tab, and then click Add.

Properties dialog box - specifying a sort:

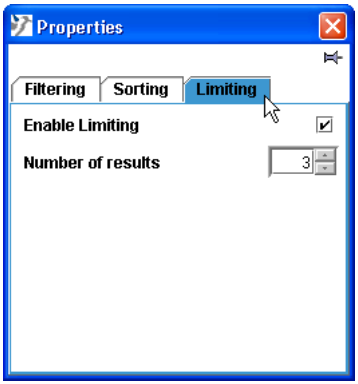


Click fields in the list, and then select to sort in either ascending or descending order.

Limiting results

To enable limiting, click the Limiting tab, and then select the Enable Limiting check box.

Properties dialog box - Limiting tab:



Use this tab to limit the number of results that are returned for proteins, peptides, and so on.

Customizing print templates

You can add pages that contain text, fields, and graphics elements (images and horizontal rules) to customize the style of the report. For example, you can add a company logo, standard company information, page numbers, and so on.

In the following examples, you will create a new page for an introduction, and add text, graphics, and fields to the header, footer, and the introduction page.

The examples illustrate the kind of objects that can be added to a template – you can insert a paragraph, field, image, or horizontal rule anywhere on any page.

Prerequisite: The following sections assume that a print template is open for modification. You can customize one of the built-in templates, or work with one you created yourself (see Creating print templates on page 11-13).

To add pages:

1. In the template navigator tree, right-click the Introduction node, and then click Insert > Page.
2. In the tree, right-click the new page, and then click Rename. Change the page name to Template Details.

When adding pages, you can display a grid, which helps you to locate and align the elements.

To display the grid:

1. In the menu bar, click View > Toggle Grid.
2. Change the size of the grid in the Preferences dialog box (see Printing tab on page 2-16).

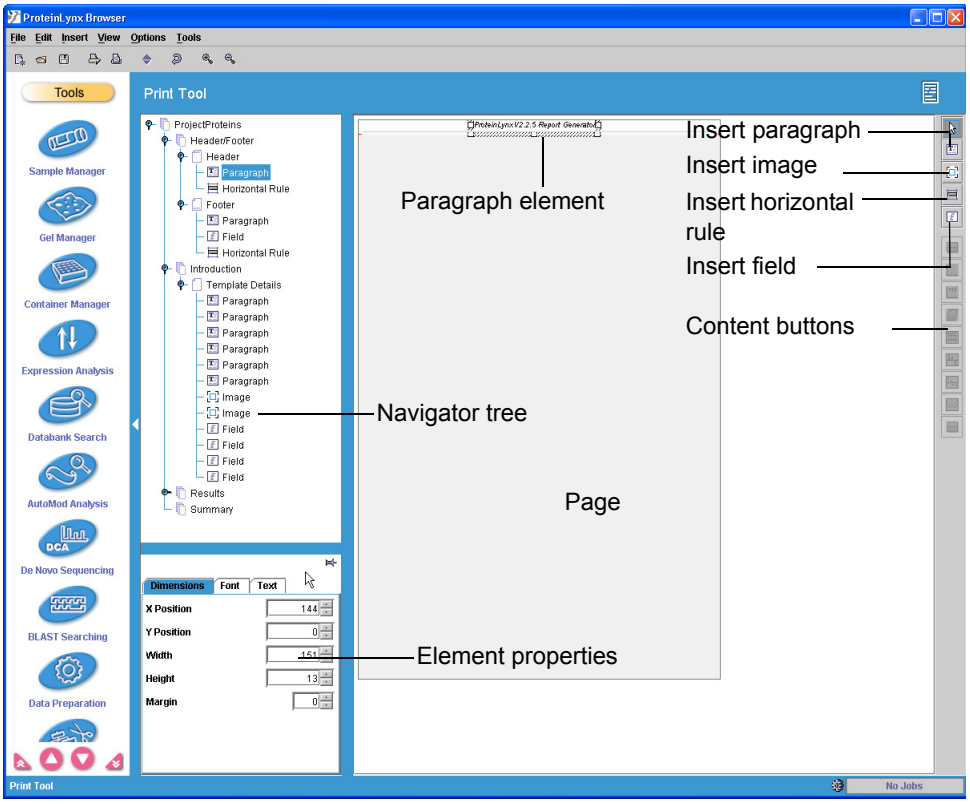
To add paragraphs:

1. In the template navigator tree, right-click the Header node.
2. Click Insert > Paragraph.

   Tip: This method inserts the paragraph in a default location and with a default size in the page. To insert a paragraph box in a location and with a size of your choice, use the buttons on the right of the browser screen.

For more details on using the buttons, see Buttons for adding content to pages on page 11-23.

Print Tool - adding paragraphs:



The four element buttons indicated are available for all types of page. Other content buttons become active depending on the type of page selected.

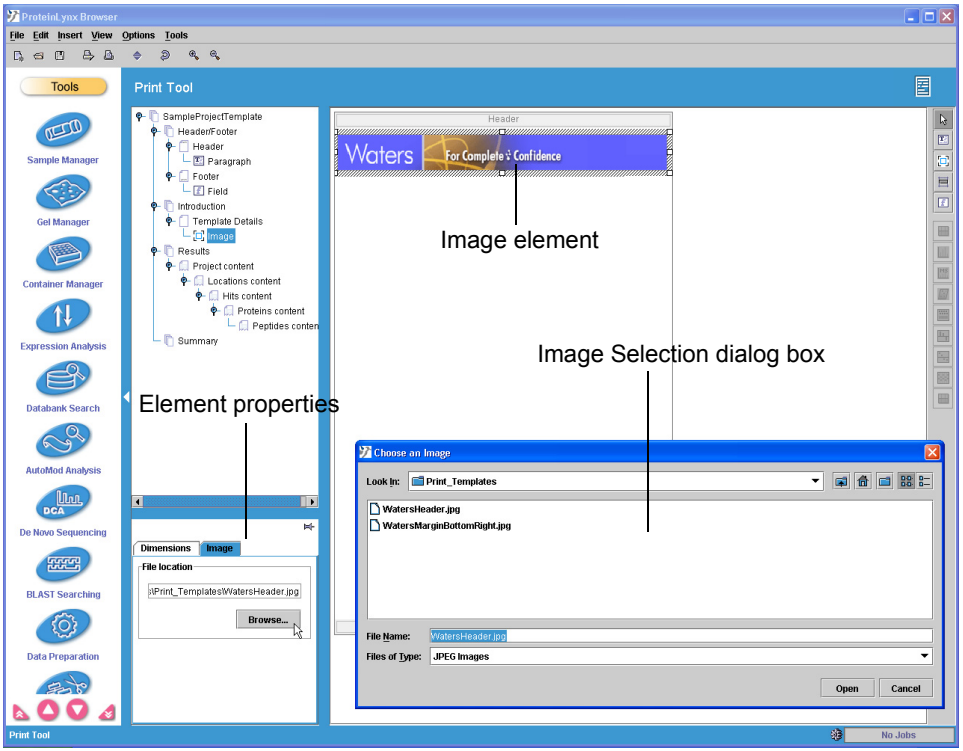
3. Add the text. You can then use the tabbed pages in the dialog box under the navigator tree to change the position, text box dimensions, font and text details.

   Tip: To center text on a page easily, size the text box to the full width of the page, and then use the Text tabbed page to set the justification to center.

To add images:

1. Right-click the Template Details node, and then click Insert > Image.
2. In the dialog box under the navigator tree, click the Image page, and then click Browse.

Print Tool - adding images:

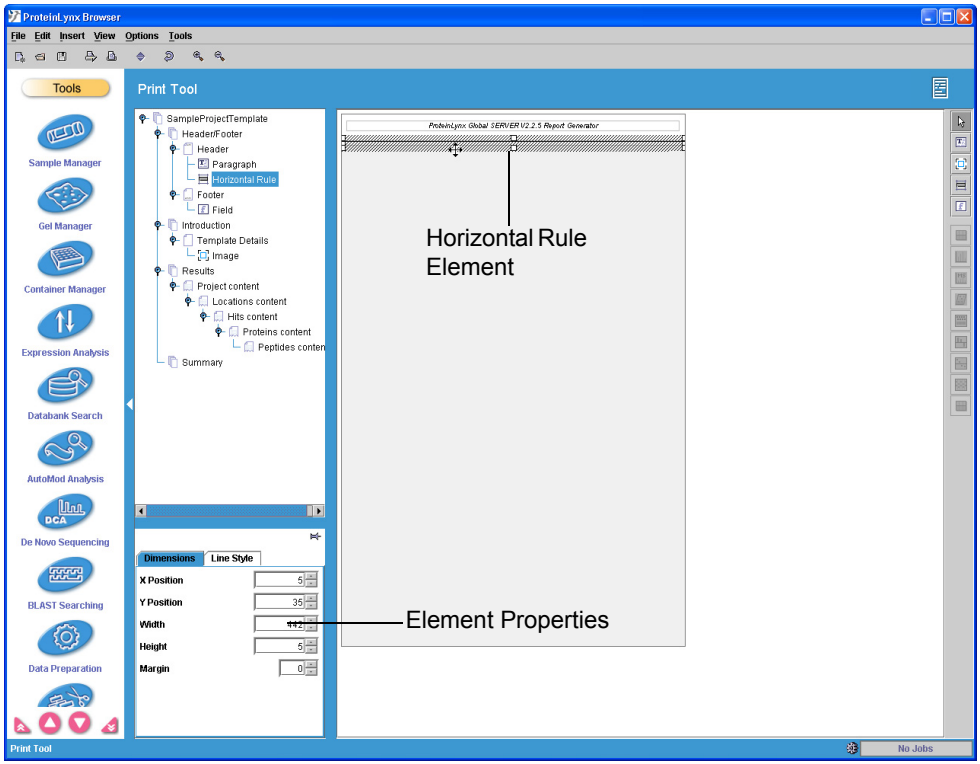


3. In the dialog box, browse to an image file, click the file, and then click Open.
4. Use the settings in the Dimensions tab to change the position and dimensions of the graphic.

To add horizontal rules

1. Right-click the Header node, and then select Insert > Horizontal Rule.

Print Tool - adding horizontal rules:

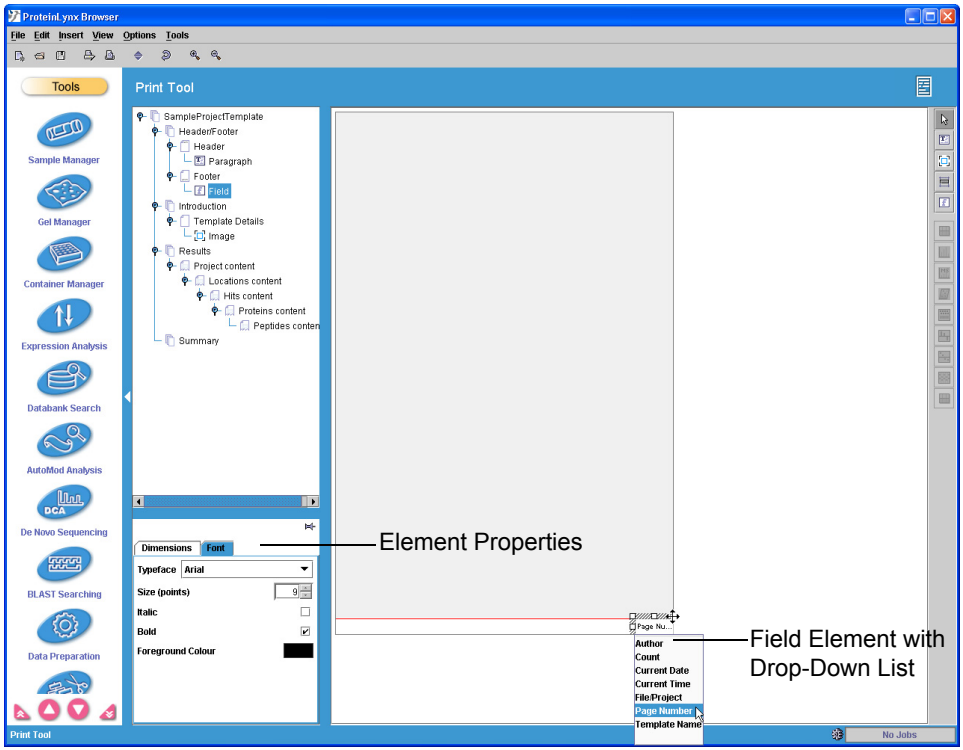


2. Use the settings in the tabs to change the line style and dimensions of the rule.

To add fields:

1. Right-click the Footer node, and then click Insert > Field.
2. In the page, click the Field box to open a drop-down list.
3. In the list, click Page Number.

Print Tool - adding fields:



4. Use the tabs to change the font and dimensions of the field.

Buttons for adding content to pages

You can use the buttons on the right of the browser to add content to a page.

The first four buttons, for adding paragraphs, images, horizontal rules and fields, are available for all pages. Whether the other buttons are available depends on the type of page selected (which controls the type of content that can be added).

The buttons enable you to drag a rectangle to the required size anywhere in a page.

The details of the buttons are shown in the following table.

Print Tool - buttons for adding content to pages:

Button	Function
	Inserts text box for a text paragraph. Available for all pages.
	Inserts an image box for a user-defined image. Available for all pages.
	Inserts a horizontal rule. Available for all pages.
	Inserts a field box for a selectable, standard, predefined field. Available for all pages.
	Inserts a box to display a table for live data. Available only for table nodes.
	Inserts a box to display an MSMS spectrum showing fragmentation data. Available only for a Peptides content page.
	Inserts a box to display an MS spectrum showing precursor data. Available only for a Proteins content page.
	Inserts a box to display a gel image showing protein separation. Available only for a Project content page.
	Inserts a box to display a coverage map showing matched peptide locations. Available only for a Proteins content page.
	Inserts a box to display an influence display showing influences. Available only for a Peptides content page.
	Inserts a box to display delta masses. Available only for a Peptides content page.
	Inserts a box to display fragment ion data. Available only for a Peptides content page.

Print Tool - buttons for adding content to pages: (Continued)

Button	Function
	Inserts a box to display workflow template parameters. Available only for a Workflow content page.

12 Managing modifier and digest reagents

Use the Modifier and Digest Reagent tools to manage the modifier and digest reagents used in the system.

Contents:

Topic	Page
Getting Started with the Modifier tool	12-2
Viewing existing modifier reagents	12-3
Adding and editing custom modifier reagents	12-4
Getting started with the Digest Reagent tool	12-7
Viewing existing digest reagents	12-8
Custom digest reagents	12-9

Getting Started with the Modifier tool

The Modifier tool enables you to manage all modifier reagents used in the ProteinLynx system. With it, you can perform these tasks:

- View the properties of the large number of modifier reagents that are supplied with ProteinLynx.
- Define your own modifier reagents, which are immediately available to the full suite of ProteinLynx browser tools.

To open the Modifier Tool, click the Modifier Tool icon on the tool tray.

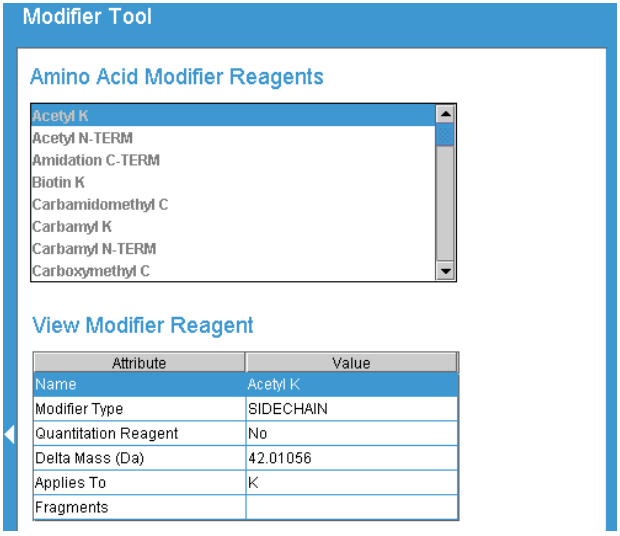
A list of modifier reagents is displayed. Supplied reagents are shown in gray text; custom, user-defined reagents are shown in black text.

Any modifier – whether supplied or custom – can be used in an isotopically-labeled experiment, so long as its Quantitation Reagent attribute is set to Isotopic.

Viewing existing modifier reagents

To view the properties of a reagent, click a reagent in the list. The attributes and values are displayed in the panel below the list.

Modifier Tool - existing modifier reagents lists:



See Reagent attributes: on page 12-4 for details of the attributes and values.

Rule: The values of supplied modifier reagents (gray text) cannot be edited.

Adding and editing custom modifier reagents

To add or edit a custom modifier reagent:

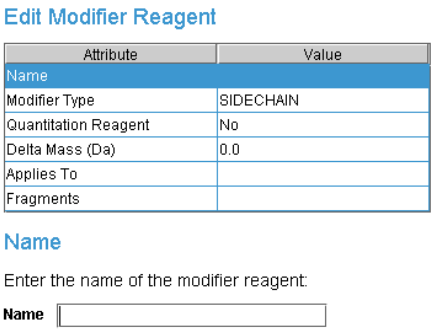
1.

- To add a reagent, click the New button on the toolbar, or click File > New.
- To edit an existing custom reagent, click the reagent in the list.

Tip: Existing custom modifier reagents are shown in black text in the list.

For both actions a panel and text box are generated, which enable defining or editing of the values for each attribute.

Adding a new modifier reagent:



Rule: Only user-defined modifier reagents can be edited; you cannot edit the supplied modifier reagents.

2. Click a row in the panel to update the value of the attribute. You can amend the values for the following attributes.

Reagent attributes:

Attribute	Description
Name	Type a unique, descriptive name; this name is used throughout the system. The supplied reagents use the format: <reagent name> <residues or terminus>.

Reagent attributes: (Continued)

Attribute	Description
Modifier type	A modifier applies to one of three 'sites' of a protein: the SIDECHAIN, N-TERM or C-TERM. Choose one from the drop-down list. If a modifier can apply to both sidechain residues and termini, define a different reagent for each case.
Quantitation Reagent	Whether this reagent should be considered a quantitation labeling reagent to be used in isotopic (ICAT, for example) or isobaric (iTRAQ, for example) labeling experiments: **Rule:** To be considered the reagent must have a positive delta mass.
Delta Mass	Delta mass is the mass difference of an amino acid residue after it has been modified by the reagent being specified.
Applies to	This attribute represents the amino acid(s) that this particular modifier can apply to. In the case of reagents applying to sidechains, these represent the modified residues themselves. For terminus modifications, any reagents specified will limit the modification to termini with an appropriate residue at the terminus. An example of this is pyrrolidone carboxylic acid N-TERM, which can only occur on N-termini adjacent to a glutamine.
Fragments	The space-separated masses and probabilities of any fragment ions resulting from this modifier reagent.

3. To save the new or edited modifier reagent, click the Save button .

   Result: The new reagent is added to the list in black text.

Deleting custom modifier reagents

To delete a custom modifier reagent, click the reagent in the list, and then either:

- Click File > Delete.
- Click the Delete button, .

Getting started with the Digest Reagent tool

The Digest Reagent Tool enables you to manage all digest reagents used in the ProteinLynx system. You can:

- View the properties of the large number of digest reagents that are supplied with ProteinLynx.
- Define your own digest reagents, which are immediately available to the full suite of ProteinLynx browser tools.

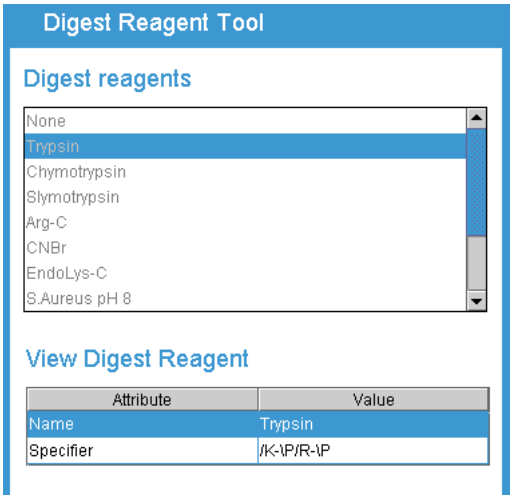
To open the Digest Reagent Tool, click the Digest Reagent Tool icon on the tool tray.

A list of digest reagents is displayed. Supplied reagents are shown in gray text; custom, user-defined reagents are shown in black text.

Viewing existing digest reagents

To view the properties of a reagent, click a reagent in the list. The attributes and values are displayed in the panel below the list.

Digest Reagent Tool:



See New digest reagent attributes: on page 12-9 for details of the attributes and values.

Rule: The values of supplied digest reagents (gray text) cannot be edited.

Custom digest reagents

You can add, edit, save and delete custom digest reagents.

Adding or editing custom digest reagents

To add or edit a custom digest reagent:

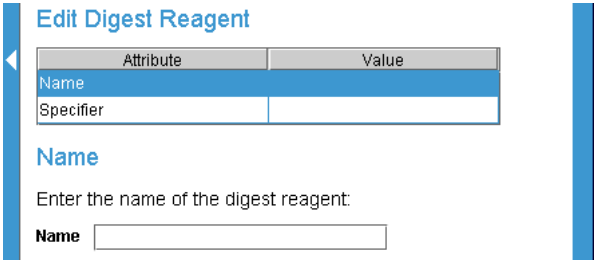
1.

- To add a reagent, click the New button , or click File > New.
- To edit an existing custom reagent, click the reagent in the list.

Rule: Existing custom digest reagents are shown in black text in the list.

For both actions a panel and text box are generated, which enable defining or editing of the values for each attribute.

Adding a new digest reagent:



Rule: Only user-defined digest reagents can be edited; you cannot edit the supplied reagents.

2. Click a row in the panel to update the value of the attribute. You can amend the values for the following attributes.

New digest reagent attributes:

Attribute	Description
Name	Type a unique, descriptive name.

New digest reagent attributes: (Continued)

Attribute	Description
Specifier	Edit this attribute to specify the cleavage points and exclusions of this reagent: The syntax of the specifier is as follows: • / forward slash indicates a cleavage point. • \ back slash indicates an exclusion for that cleavage, for the C-terminus. • -\ hyphen then back slash indicates an exclusion for that cleavage, for the N-terminus

Saving custom digest reagents

To save the new or edited digest reagent, click the Save button, .

Result: The new reagent is added to the list in black text.

Deleting custom digest reagents

To delete a custom digest reagent, select the reagent in the list, and then either:

- Click File > Delete.
- Click the Delete button, .

13 Organizing databanks with the Databank Admin tool

Contents:

Topic	Page
Getting started with the Databank Admin tool	13-2
Adding databanks	13-3
Editing databanks	13-11
Removing and deleting databanks	13-13
Connecting to a search engine	13-17

Getting started with the Databank Admin tool

Databanks are flat files that contain information regarding sequences of nucleotides or amino acids. These files are used by the Databank Search and the BLAST Searching tools.

The Databank Admin Tool:

- Enables you to organize databanks and choose databank properties.
- Regulates any automatic downloads and updates.
- Generates auxiliary files that are needed by the other tools when performing searches.
- Enables you to view the databanks that reside on the currently connected search engine.

To open the Databank Admin Tool, click the Databank Admin Tool icon in the tool tray.

Tips:

- A search engine must be specified (see Changing preferences on page 2-5) for the Databank Admin options to be available.
- If there are no databanks displayed when the Databank Admin Tool opens, try restarting the search engine. For help with starting modules, see Chapter 1 - Installing ProteinLynx Global SERVER.

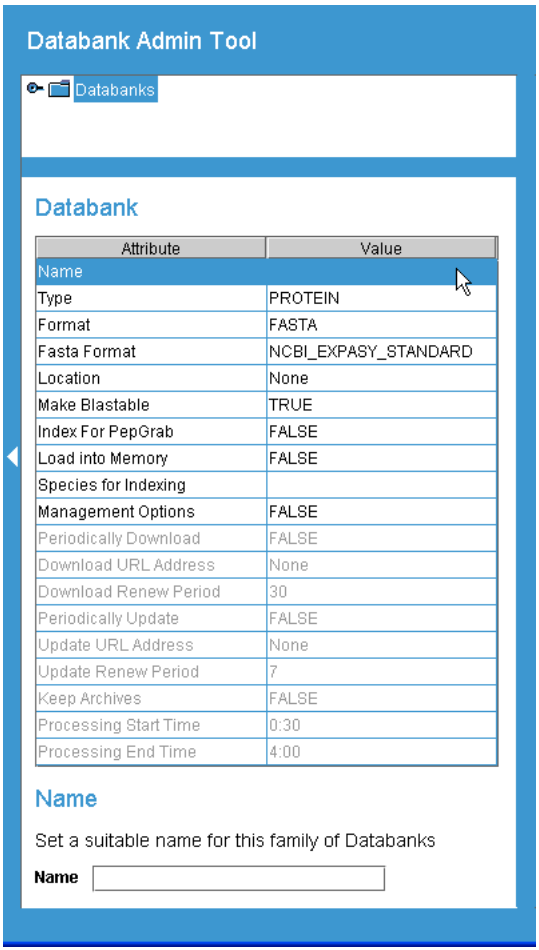
Adding databanks

To add a new databank:

1. Click on the toolbar, or click File > New Databank.

   A Databank editor panel opens under the navigator tree.

   #### Databank editor panel and navigator tree:



2. To change the values of any attributes, click the attribute in the panel, and then edit the value under the panel. See Databank attributes on page 13-4 for details of the attributes and values.

3. Click to save the new databank.

   The new databank is displayed in the navigator tree.

   If the file is large, processing of the databank file can take several seconds. When the file has been processed, the databank is available to the various Protein Probe tools and the databank name is displayed in the Databanks field of the Databank Search Tool.

To ensure that the most up-to-date state of the Databanks are being displayed in the Databank Admin Tool, click (Refresh Databanks Tree View) on the toolbar.

Databank attributes

You can change the values for the following attributes:

Databank attributes:

Attribute	Description
Name	Contains the name of the family of databanks. This name appears in the list of databanks in the Databank Search tool and other search tools. This field is compulsory, and must be set when a new databank is created. After the databank has been created and saved, this field cannot be changed.
Type	Select from the list of supported databank types. Default = Protein.
Format	The format of the sequences in the databank flat file. Select from the list of supported formats. It is important that the correct format is selected so that the databank can be processed correctly and that search results can be displayed in a meaningful way.

Databank attributes: (Continued)

Attribute	Description
FASTA Format	One of the most widely used formats for specifying sequence information is FASTA format. In its most general form, FASTA format comprises a one line description beginning with a '>' symbol followed by multiple lines containing the sequence of amino acid identifiers. Within this general format, there are many format subtypes used by different organizations. If the format of the databank is FASTA, use this field to specify the particular FASTA convention which is used. From the list of supported FASTA formats, select whichever subtype corresponds to the sequences in the flat file. Formats are: STANDARD NCBI_EXPASY_STANDARD NCBI_PRF_PIR NCBI_PDB NCBI_PATENT NCBI_GENINFO NCBI_GENERAL NCBI_LOCAL PDB PIR SRS ARABIDOPSIS_GENOME NRDB UNIGENE STANDARD_SPACED LONG_DESCRIPTION ACCESSION_ONLY UNKNOWN If the format is not FASTA, this field is ignored. **Requirement:** In order for search results to contain accession numbers, and therefore be suitable for protein quantification in Expression Analysis, the FASTA format must be set correctly. **See also:** For definitions of the FASTA formats, see FASTA flat file format on page E-9.

Databank attributes: (Continued)

Attribute	Description
Location	This field is compulsory. Enter the file path of the flat file where the databank flat file is located. When a databank has been created and saved, this field cannot be changed. If there is already a flat file of sequences for the databank, use the File dialog box to choose this file. If there is no flat file yet in existence for this databank, and if the databank will be automatically downloaded, choose the location to which the databank should be downloaded. **Requirement:** If the databank resides outside of the PLGS installation directory, the Windows users who will run PLGS must have read, write, and modify access to the databank directory. This requirement is especially relevant if the user adding the databank is an administrator and the users running PLGS are not.
Make Blastable	If this option is set to TRUE, the necessary index files are created and the databank will be available for BLAST searching by using the BLAST Searching tool.
Index For PepGrab	If this option is set to TRUE, the necessary index files are created and the databank will be available for PepGrab searching via the PepGrab function.
Load into Memory	Loading a databank into memory increases the speed at which that databank can be searched by the Databank Search Tool. Ensure sufficient RAM is available. Select True or False as required. **Tip:** Databank searches can fail if very large disk-based databanks are used. If a failure occurs, try loading the databank into memory.

Databank attributes: (Continued)

Attribute	Description
Species for Indexing	When a databank has been indexed by a species, a Databank Search restricted to that species can be performed using the Databank Search Tool. Any number of species for indexing can be selected for indexing. Each species for which the databank has been indexed will appear in the Databank Search Tool species list for that databank. Select any combination of species. To select more than one species, hold down the Ctrl key while clicking the required list elements.
Management Options	If further management options are required, set this option to TRUE. This will make available further options relating to automatic downloads, automatic updates and keeping of archives. Select True or False as required. **Requirement:** When a new version of a databank is downloaded, any workflow templates that relate to the databank must be updated with the new version number.
Periodically Download	To periodically download the databank from a remote location, set this option to TRUE. **Rule:** This attribute is only available if the Management Options attribute is set to TRUE. If this attribute is set to True, you must specify a remote location URL from which the databank will be periodically downloaded (Download URL Address field). There are several other options relating to periodic downloading that can be set or be left at their default values. These are: • Download Compression Type • Download Renew Period • Keep Archives • Processing Start Time • Processing End Time

Databank attributes: (Continued)

Attribute	Description
Download URL Address	**Rule:** This option is only available if the Periodically Download attribute is set to True. You must set this if the Periodically Download attribute has been set to True. This field contains the URL address from which the databank should be periodically downloaded. 1. Click the URL button, and then type the URL address in the URL field. 2. Click Open on the URL Chooser. The system locates the remote address and checks that it can be accessed. This can take a few seconds.
Download Compression Type	**Rule:** This option is only available if the Periodically Download attribute is set to True. This field relates to the periodic download of remote files. Databank flat files available at public sites are often stored in a compressed form to save space. The Databank Admin tool will automatically decompress several types of compressed file, including .z .Z .zip and .gz compression types. If known, you can specify the compression type of the remote file. If the field is left as Unknown then the system decides the compression type.
Download Renew Period	**Rule:** This option is only available if the Periodically Download attribute is set to True. Enter the number of days after which a new databank flat file will be downloaded. Download processing will only take place between the Start and End times. The default period between downloads is 30 days. In the text box, type a whole number greater than zero.

Databank attributes: (Continued)

Attribute	Description
Periodically Update	**Rule:** This option is only available if the Management Options attribute is set to True. To periodically update the databank from a remote location using interim update files, set this option to True. Some providers of databanks supply interim update files, which contain only recently added sequences. Performing updates reduces the need for frequent full downloads of databanks, which can use a lot of resources. If this attribute is set to true, you must set the Update URL Address attribute to specify a remote location URL from which the databank will be periodically updated. There are several other options relating to periodic updating that can be set or be left at their default values: • Update Compression Type • Update Renew Period • Keep Archives • Processing Start Time • Processing End Time
Update URL Address	This field must be set if the Periodically Update attribute has been set to True. This field contains the URL address from which the databank should be periodically updated. 1. Click the URL button to open the URL Chooser dialog box. 2. Type the URL address of the remote file from which the databank should be periodically updated, and then click Open. The system locates the remote address and checks that it can be accessed. This can take a few seconds.
Update Compression Type	**Rule:** This option is only available if the Periodically Update attribute has been set to True. The details of this attribute are the same as for Download Compression Type.

Databank attributes: (Continued)

Attribute	Description
Update Renew Period	**Rule:** This option is only available if the Periodically Update attribute has been set to True. Enter the number of days after which an automatic interim update will be undertaken. The details of this attribute are the same as for Download Renew Period.
Keep Archives	**Rule:** This option is only available if one or both of the Periodically Download or Periodically Update attributes have been set to True. To keep archived databanks, set this field to True. These archives can be restored at a later date. For details of archives, see Keeping archived copies of a databank on page 13-15).
Processing Start Time and Processing End Time	Format: HH:MM (24-hour clock). Some of the processing steps, such as automatic download of large databank files and making blastable, when applied to large files, can take time to perform. During this processing period, the databank might become temporarily unavailable to other search tools. For this reason, it can be preferable to schedule processing to take place only at times when the databanks are unlikely to be needed by other tools. The Processing Start Time specifies the time after which all such automatic processing will be scheduled. The Processing End Time specifies the time after which no further processing will be scheduled. It is important to specify a time period during which the machine is on. If there is no preferred processing time, set to 00:01 and 23:59.

Editing databanks

You can only edit databanks that reside on the local machine and are administered by the local search engine.

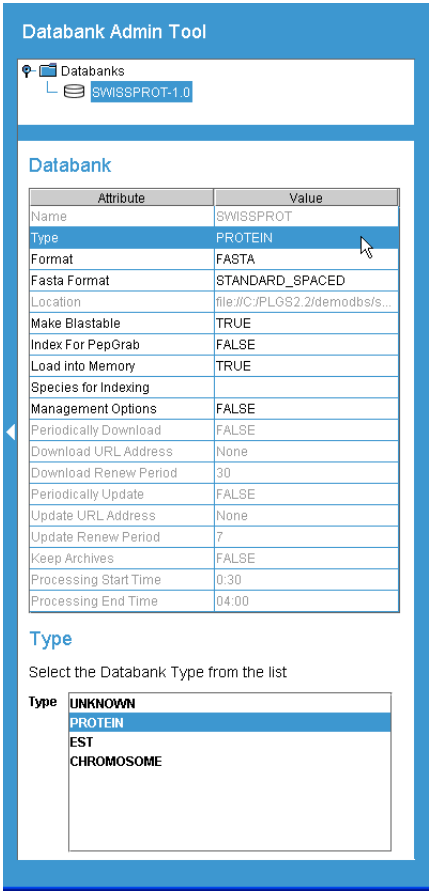
Databanks that reside on a remote machine and are administered by a remote search engine can only be viewed, not edited.

To edit a databank:

1. Click the databank in the navigator tree.

   The Databank Editor Panel is displayed.

Databank Editor Panel:



2. Click the required attributes in the panel, and then edit them at the bottom of the panel. For details of the attributes, see Databank attributes on page 13-4.

   Rule: You cannot edit values for the Name or Location attributes.

3. Click the Save button to save the databank.

Removing and deleting databanks

If a databank resides on the local machine, you can:

- Remove the databank, but not its associated files. A removed databank can be revived (restored) later.
- Delete the databank, including its associated files. A deleted databank cannot be revived (restored) later.

Removing databanks from the system record

Using the Databank Admin Tool, you can remove any databank that resides on the local computer.

To remove the databank from the machine, but not remove the files associated with the databank:

1. In the navigator tree, click the databank to be removed.
2. Click the Remove button on the toolbar.
3. Confirm the request when prompted.

Results:

- The databank is removed from the record of the Databank Admin Tool.
- The databank will no longer appear in the navigator tree and will not be available for searching by the various ProteinLynx tools.
- The files associated to the databank, including the flat file of sequences, will not be removed from the computer.

Deleting databanks

Using the Databank Admin Tool, you can delete any databank that resides on the local computer.

To delete a databank from the machine, including the files associated with the databank:

1. In the navigator tree, click the databank to be deleted.
2. Click the Delete button on the toolbar.

3. Confirm the request when prompted.

Results:

- The databank is removed from the record.
- The files associated to the databank, including the flat file of sequences are deleted from the computer.
- Any auxiliary files used for BLAST searching, and any archive files are also deleted.
- The databank no longer appears in the navigator tree and is not available for searching.

Deleting archive files

You can delete the archives of any databanks that reside on the local computer.

To delete an archive without deleting the entire databank:

1. In the navigator tree, expand the node of the databank.
2. Click the node of the archive which is to be deleted.
3. Click the Delete button on the toolbar.
4. Confirm the request when prompted.

Results:

- The archive is removed from the Databank Admin Tool record and no longer appears in the navigator tree.
- The underlying zipped archive file is deleted from the file system.
- The archive is not available for future revival.

Deleting revived archives

You can delete revived archives which reside on the local computer.

To delete a previously revived archive without deleting the entire parent databank:

1. In the navigator tree, expand the node of the relevant databank.

2. Click the node of the revived archive which is to be deleted.

Revived archive node (dark colored)

Archive node (grayed-out)

3. Click the Delete button on the toolbar.
4. Confirm the request when prompted.

Results:

- Any files needed for search processing are deleted from the file system.
- In the navigator tree, the color of the node changes to gray to indicate an archived databank.
- The corresponding version of the databank is not available to the various search tools.
- The zipped archive file remains in the file system. The archive is still available for revival in the future.

Keeping archived copies of a databank

Databanks can change over time as new sequences are added, or the databank is periodically downloaded or updated. Therefore, archived copies of databanks are useful, especially if you want to repeat previous experiments using the original databank.

To keep archives of databanks, set the Keep Archives attribute to True when creating or editing a databank (see Adding databanks on page 13-3 and Editing databanks on page 13-11). This creates a zipped (compressed) file of the databank. However, you must also consider that large databanks create large zipped archive files. Therefore, consider whether your system has sufficient resources available to store archives.

Reviving an archive

If archives exist for a databank that resides on the local computer, these archives can be revived (restored) for use by the various search tools. For example, you might want to revive an archive to verify results that were obtained from an older version of the databank.

To revive an archive:

1. In the navigator tree, expand the node of the relevant databank.
   The available archives appear as gray-colored icons .
2. Click the archive to be restored.
3. Click the Revive button on the toolbar.
4. Confirm the request when prompted.
   The color of the node changes , which indicates that the archive has been restored.

The corresponding version of the Databank is available for searching by the Databank Search tool and, if appropriate, the BLAST Searching tool. The databank version appears in the list of searchable Databanks for each of those tools.

Connecting to a search engine

The ProteinLynx browser interface communicates with the ProteinLynx Search Engine, which regulates Databank searches, AutoMod searches, De Novo searches and BLAST searches. The Search Engine can be present on the local machine. Alternatively, ProteinLynx browser can be connected to a Search Engine residing on a remote machine.

Connect to an alternate Search Engine by using the Preferences button and dialog box (see Changing preferences on page 2-5). When the procedure has been completed, ProteinLynx will connect to the Search Engine on the machine specified.

Rule: Databanks which reside on the local machine and are administered by the local search engine can be viewed, searched and edited. Databanks which reside on a remote machine and which are administered by a remote Search Engine can be viewed and searched but cannot be edited.

14 Query Tools

This chapter outlines the query tools that are available within ProteinLynx Global SERVER.

By default, these tools are not displayed in the tool tray or Tools menu. To add the tools, follow the instructions in Adding and removing tools on page 2-4.

- Databank Search tool – Enables you to search both MS and MSMS spectra data against a selected databank to identify the protein(s) contained in the original sample.
- AutoMod Analysis tool – Increases protein coverage and reduces unmatched MSMS spectra by taking the protein sequences identified through databank searching and rigorously analyzing these against the submitted spectra.
- De Novo Sequencing tool – Enables you to determine the primary sequence of a peptide directly from its MSMS data.
- BLAST (Basic Local Alignment Search Tool) Searching tool – Performs a homology search on the selected databank using the input protein/peptide sequences.

Use these tools to create and edit individual queries, submit those queries to the search engine, and view the query results.

Contents:

Topic	Page
Databank Search tool	14-3
AutoMod Analysis tool	14-14
De Novo Sequencing tool	14-19
BLAST Searching tool	14-23

Query toolbar

All the query tools share the same toolbar buttons:

Query toolbar buttons:

Button	Description
	Submits the current query to the search engine.
	View and edit preferences.

Databank Search tool

The Databank Search tool enables you to search spectrum data against a protein or EST databank that has undergone a theoretical digest. This search enables you to identify the protein(s) contained in the original sample.

You can perform the following types of databank search:

- PMF (Peptide Mass Fingerprint)
- PMF + Fragmentation Ion Search
- Fragment Ion Search

Using this tool, the search type performed is dictated by the type of mass spectrum data attached.

Databank Search details:

Type of Databank Search	Type of Mass Spectrum Data
PMF	Maldi MS, or Maldi Q-Tof MS
PMF + Fragment Ion Search	Maldi Q-Tof MSMS
Fragment Ion Search	Electrospray Q-Tof MSMS, or Maldi PSD

However, using the Workflow Designer, you can generate workflow templates that allow any type of Databank Search to be applied to any type of Mass Spectrum Data.

Example: Use a PMF search for Electrospray Q-Tof MSMS data.

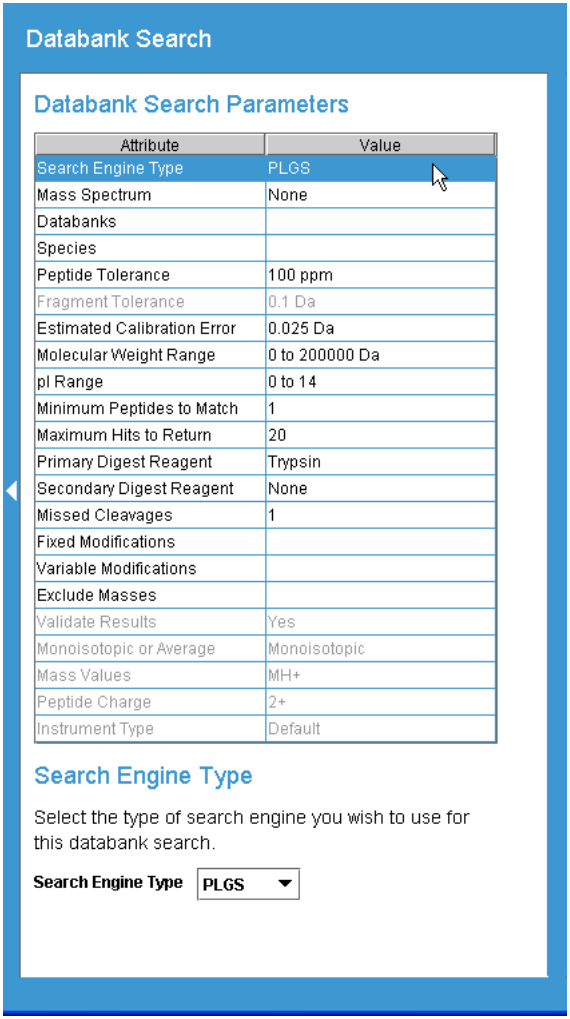
Also, using the Workflow Designer, a Databank Search can be incorporated into a workflow as the first step in a more comprehensive analysis (see Chapter 7 - Defining templates for searching with Workflow Designer).

Databank searches can be submitted not only to the ProteinLynx search engine, but also to a Mascot (version 2.0 or later) search engine. The results can be displayed in the ProteinLynx browser or an Internet browser.

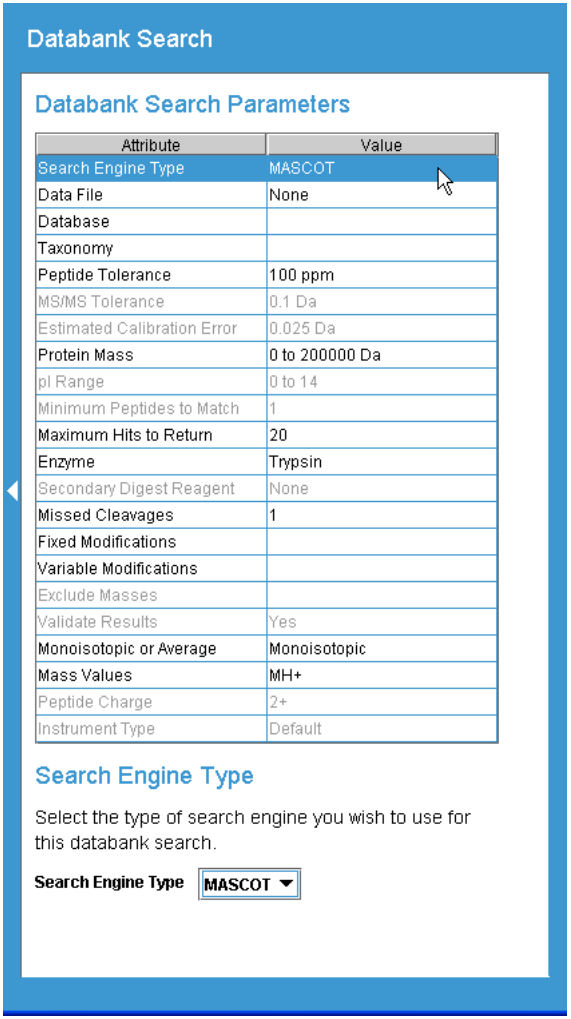
To open the Databank Search tool, click the Databank Search icon in the tool tray.

The Databank Search Parameters table opens in the Editor Panel of the browser, with the Search Engine Type attribute highlighted. The MASCOT option is available only if you have a valid connection to a Mascot search engine. For details of how to connect to a Mascot search engine using the Preferences dialog box, see Search Engine tab on page 2-5.

Databank Search parameters - for PLGS or Mascot search engines:



PLGS attributes



MASCOT attributes

To perform a Databank search:

1. Click an attribute in the table (see Databank search parameters on page 14-5 for details), and then edit the value in the panel at the bottom of the table.
2. When the required fields have been edited, click the Submit button on the toolbar to start the search.

Databank search parameters

The following sections detail the attributes in the Databank Search Parameters table.

Requirement: You must specify the attribute's Search Engine Type, Mass Spectrum (PLGS) and Databanks (PLGS) or Database (MASCOT).

Search Engine Type

You can select PLGS or MASCOT. When performing a Mascot PMF search or Mascot Fragment Ion Search, select MASCOT from the drop-down list.

Mass Spectrum (PLGS) or Data File (MASCOT)

This attribute specifies the spectrum data file on which to perform the analysis. You can choose a file or URL that contains mass spectrum data.

To select a file that contains mass spectrum data click File, and then choose a mass spectrum file. The following formats are valid.

Mass Spectrum - valid data file formats:

Type of MS data	Valid formats
MS data	MS Text (*.txt), XML (*.xml), or mzData (*.mzData)
MSMS data	PKL (.*.pkl), XML (*.xml), or mzData (*.mzData)

To specify a URL, click the URL button, and then specify or select a URL in the URL Chooser dialog box (see Figure titled "URL Chooser dialog box:" on page 7-10).

Databanks (PLGS) or Database (MASCOT)

This attribute specifies the protein or EST databank/database that the mass spectrum data is to be searched against. You can add PLGS databanks using the Databank Admin Tool (see Organizing databanks with the Databank Admin tool on page 13-1). New Mascot databases can only be made available by your Mascot server administrator.

The list contains all available databanks/databases. Click the name of a databank or database to select it.

Tip: It is advisable to specify a databank that contains the majority of protein sequences that could be in the sample data searched.

Rule: Only one databank/database can be searched at any one time; any new selection replaces the existing selection.

Species (PLGS) or Taxonomy (MASCOT)

These attributes are optional. By default, the entire databank/database will be searched for matches to the data, and all matches will be considered regardless of species or taxonomy.

PLGS databanks can be indexed according to species using the Databank Admin Tool (see Organizing databanks with the Databank Admin tool on page 13-1), which allows searches using an indexed databank to be limited to one or more species. Mascot taxonomies can only be changed by the Mascot server administrator.

To restrict the search to one or more species, click the species in the list. To select multiple species in the list, use Shift+click to select consecutive species, or Ctrl+click to select non-consecutive species.

Peptide Tolerance

This attribute is optional as a default value is supplied.

This attribute is used to match intact peptide masses. The units used for PLGS searches are parts per million (ppm) or Daltons (Da). Mascot searches have additional units available: percentage (%) and absolute millimass units (mmu). The peptide tolerance should reflect the known accuracy of the instrument used to acquire the spectrum data. Restricting this attribute to the lowest feasible value can greatly reduce search times and increase the quality of the results.

To specify the tolerance, type the value into the text field, and then click the desired units in the drop-down list.

Fragment Tolerance (PLGS) or MSMS Tolerance (MASCOT)

Restricting fragment tolerance is encouraged as it can reduce search times. Specifying a fragment tolerance is optional as a default tolerance is supplied.

Rule: This attribute cannot be modified for PMF searches, as fragmentation spectra are ignored.

This attribute is used in the final validation of Fragment Ion Search results. If the Validate Results attribute is used (see Validate Results on page 14-12), this value determines which y-ions have been matched successfully. It is recommended that this value is set to the lowest value possible, but should be at least double the value of the Estimated Calibration Error (see Estimated Calibration Error (Da or ppm) on page 14-7). This increases the quality of the validated peptide returned.

To specify the tolerance, type the value into the text field.

Estimated Calibration Error (Da or ppm)

Restriction: This attribute is not available for Mascot database searches.

The Estimated Calibration Error is an estimation of the error introduced following instrument calibration. This value is fundamental to the scoring of a peptide sequence against a given fragmentation spectrum.

As a tight error will significantly reward well-measured data in the scoring, it is recommended that spectra submitted are well mass measured, to allow a low Estimated Calibration Error to be set. It is not necessary to adjust the estimated calibration error for small variations of this number in the fourth decimal place.

When comparing calculated peptide or fragment masses with the data, it is important to know how well the masses in the data are determined. If this estimate is good, the information that can be extracted from the data is maximized. A good estimate will increase the scores of correct identifications.

Suitable values differ between instruments. Recommended values are:

Estimated Calibration Error - recommended values:

Instrument Detail	Estimated Calibration Error - recommended value
Equipped with nano-lockspray	20 ppm
MALDI equipped with internal lockmass	30 ppm
MALDI equipped with external lockmass	50 ppm

Molecular Weight Range (PLGS) or Protein Mass (MASCOT)

These attributes are optional as a default range is supplied.

Restriction: This cannot be used for searches of EST databanks.

This attribute restricts the number of returned protein matches to a range of molecular weights (PLGS) or masses (MASCOT). Specify a narrow range to reduce search times.

Tip: The range could be based on the location of the gel from which the sample that generated the data originated. When looking for a specific protein of interest, the size and range indicates the confidence in the estimation of the molecular weight or protein mass.

For a PLGS search, type the minimum and maximum molecular weights in Daltons. For a Mascot search, specify the maximum protein mass in Daltons.

pI Range

This attribute is optional: by default all proteins are searched.

This attribute restricts the number of returned protein matches to within a specific iso-electric point range. The range could be based on the location of the gel from which the sample that generated the data originated, or the range of a specific protein of interest. Using a narrow range reduces search times.

Restriction: This attribute cannot be used for searches of EST databanks or for Mascot searches.

To specify the range, type the minimum and maximum iso-electric points in the text fields.

Minimum Peptides to Match

This attribute is optional as a default value is supplied.

Rule: This attribute applies only to PLGS PMF searches.

This attribute specifies the number of peptides that have to be matched to a sequence before that sequence is considered to be a significant hit. The greater the number of matches required for a hit to be returned, the more reliable the search results will be. However, if the spectrum is of poor quality, specifying a high value could discount significant sequences.

In the text field, type the minimum number of peptides that a protein must match before it is included in the search results.

Maximum Hits to Return

This attribute is optional as a default value of 20 is supplied.

Use this attribute to specify the maximum number of hits to be included in the search results. It is recommended that you use the default value for a PLGS search of Q-Tof MSMS data.

In the text field, type the required number. If the search identifies more than the specified number of hits, only the top-scoring hits are reported.

Primary Digest Reagent (PLGS) or Enzyme (MASCOT)

This attribute is optional as a default reagent is supplied.

The list contains all available digest reagents. Click the name of a reagent to select it.

Rule: Only one reagent can be searched at any one time: any new selection replaces the existing selection.

Selecting None or Non-specific

In addition to a number of pre-defined reagents, the PLGS menu contains the options None and Non-specific. None is a suitable choice for Fragment Ion databank searches containing peptide sequences, as it means that the sequences are not digested.

Non-specific will digest sequences non-specifically, resulting in longer databank search times. This is a suitable choice for all databank search types (PMF, Fragment Ion search, and so on), although a non-specific digest can be more suited to AutoMod analysis (see AutoMod Analysis

tool on page 14-14), where a small subset of databank entries can be submitted for characterization.

A non-specific digest reagent generates all the possible peptides, up to a length of 30 amino acids, for each databank entry. It is recommended that you do not select a non-specific reagent without the use of additional filters, due to the large number of theoretical peptides that will be produced.

Rule: If an AutoMod search is part of a search sequence, and a Non-specific digest reagent is specified, all proteins will show 100% missed cleavages, irrespective of which digest reagent was used in the preceding databank search step.

For a PLGS search, to add alternative reagents to the existing list, use the Digest Reagent Tool (see Getting started with the Digest Reagent tool on page 12-7). For Mascot searches, see your Mascot server administrator.

Secondary Digest Reagent

This attribute is optional as a default reagent is supplied.

If two digest reagents are applied to a sample, they are applied sequentially. Therefore, a theoretical digest using a second reagent is carried out on peptides produced by the first digest.

Select a reagent from the list, as for Primary Digest Reagent (PLGS) or Enzyme (MASCOT) on page 14-9.

Missed Cleavages

This attribute is optional as a default number is supplied.

This attribute specifies the maximum number of missed cleavages permitted when generating the set of peptides produced by a theoretical protein digest. The value is applied to the primary and secondary digest reagents, except where a non-specific reagent or None is selected.

Fixed Modifications

This attribute is optional. By default, no Fixed Modifications will be applied to the peptides produced by the digests.

The list contains all available modifier reagents.

To specify a modification that should always be applied to peptides produced by the digests, click the desired reagent in the list. To select multiple reagents in the list, use Shift+click to select consecutive reagents, or Ctrl+click to select non-consecutive reagents.

For a PLGS search, to create additional modifiers to the existing list, use the Modifier Tool (see Getting Started with the Modifier tool on page 12-2). For Mascot searches, see your Mascot server administrator.

Variable Modifications

This attribute is optional. By default, no variable modifications are applied to the peptides produced by the digest.

You can apply any number of variable modifications to the peptides generated by the theoretical digest. However, if search times are critical, you need to consider carefully the use of this attribute.

Example: If a single variable modification is applied, a peptide containing three amino acids that bond with the modifier will generate eight variations in Fragment Ion searches and four in PMFs.

To specify a modification that should always be applied to peptides produced by the digests, click the desired reagent in the list. To select multiple reagents in the list, use Shift+click to select consecutive reagents, or Ctrl+click to select non-consecutive reagents.

Exclude Masses

Rule: This attribute applies only to PLGS PMF searches.

This attribute specifies masses that are to be excluded from a search. These excluded masses could include masses of known matrix impurities, contaminants, or lockmass peaks. If the specified masses appear in the submitted spectra to within the supplied peptide tolerance, these masses are suppressed when performing the search. The masses are not actually excluded, but their influence is suppressed as it is assumed that the peaks belong to a contaminant. Therefore, while excluded masses can sometimes be matched, the influence that these peaks contribute to the final score is suppressed.

In the text box, type the masses that are to be excluded, separated by a space, or return (MALDI only).

Masses selected for exclusion are usually theoretical masses, which can differ from masses found in the data. Therefore, due to the possibility of mis-assignment, the corresponding data is suppressed according to how well the masses match the theoretical masses rather than being completely extinguished.

Validate Results

All MSMS results can be validated. A validated peptide will contain a series of three or more consecutive y-ions.

If validation is selected, the top scoring peptide for each MSMS spectrum is returned. This could increase the requirement for manual validation of the results returned.

To validate the results, select the check box.

Monoisotopic or Average

Rule: This attribute applies only to Mascot searches.

This attribute specifies whether the mass values used in the search are monoisotopic or average. In the drop-down list, click:

- Monoisotopic – mass of the first peak in an isotope distribution.
- Average – centroid of the whole isotope distribution.

Mass Values

Rule: This attribute applies only to Mascot PMF searches.

This attribute specifies whether the experimental peptide mass values in a PMF search include the mass of the charge-carrying proton (MH+), or if they correspond to neutral values (Mr).

Click the relevant values in the drop down list.

Peptide Charge

Rule: This attribute applies only to Mascot Fragment Ion searches.

This attribute specifies the precursor peptide charge state in a Fragment Ion Search.

Click the charge state in the drop down list.

Instrument Type

Rule: This attribute applies only to Mascot Fragment Ion searches.

This attribute specifies the instrument that was used to acquire the data, which determines the fragment ion series used for Mascot scoring.

Click the type of instrument in the drop-down list.

AutoMod Analysis tool

AutoMod increases protein coverage and reduces unmatched MSMS spectra by taking the protein sequences identified through databank searching and rigorously analyzing them against the submitted spectra.

The analysis can consist of any combination of non-specific cleavages, post-translational modifications, and amino acid substitutions. The speed of the search is as a consequence of analyzing only those sequences that have already been identified, rather than laboriously trailing through the entire databank.

Tip: Using the algorithm in automated workflows (see Chapter 7 - Defining templates for searching with Workflow Designer) can increase coverage and confidence of the top databank search hits, while simultaneously filtering out questionable, lower-scoring hits.

You can use the AutoMod Analysis tool to search data from any instrument that can generate fragmentation spectra: Electrospray Q-Tof, Maldi PSD and Maldi Q-Tof.

To open the AutoMod Analysis query tool, click the AutoMod Analysis Icon in the tool tray.

The AutoMod Search Parameters table opens in the editor panel of the browser.

AutoMod Analysis search parameters:

To perform an AutoMod Analysis search:

1. Click an attribute in the table (see AutoMod Analysis search parameters on page 14-16 for details), and then edit the value in the panel at the bottom of the table.
2. When the required fields have been edited, click the Submit button on the toolbar to start the search.

When the analysis is complete, the results are displayed in the unified results panel that is added to the desktop.

AutoMod Analysis search parameters

The following sections detail the attributes in the AutoMod Search Parameters table.

The attributes Mass Spectrum, Peptide Tolerance, Fragment Tolerance, Estimated Calibration Error, Primary Digest Reagent, Secondary Digest Reagent, Missed Cleavages and Fixed Modifications and Validate Results are described in Databank search parameters on page 14-5.

Consider Modifications

You can specify whether modifications should be considered in the matching of spectra against generated peptides. If modifications are considered (default), all the modifications listed in the Modifier Tool are considered, where appropriate.

The check box is selected by default. Clear the check box to specify that modifications should not be considered.

Consider Substitutions

You can specify whether single amino acid substitutions should be considered in the matching of spectra against generated peptides. If substitutions are considered (default), all the substitutions listed in the Modifier Tool are considered, where appropriate.

The check box is selected by default. Clear the check box to specify that substitutions should not be considered.

Specify which substitutions to consider in the Substitution Likelihood attribute (see Specifying the likelihood of substitutions on page 14-17).

Specifying the maximum substitutions and modifications per peptide

In the Max. Mods/Subs per Peptide attribute you must specify a maximum number of modifications and/or substitutions to be considered per starting peptide.

This figure limits the number of residues per peptide that can be modified or substituted at any one time.

> **Example:** Consider the case after digestion that the following starting peptide is generated:
>
> ACDEFGHILK (10 residues)

Now, consider that only substitutions are being considered (no modifications) and that all substitutions are valid. Each residue can therefore undergo 19 different substitutions.

Considering a maximum of 0 mods/subs per peptide will generate only 1 peptide: the starting peptide above.

Setting max. mods/subs to 1 will generate 191 ((10 x 19) + 1) potential matching peptides.

Considering a maximum of 2 mods/subs per peptide will now generate 16436 ((45 x 19 x 19) + (10 x 19) + 1) potential matching peptides.

Therefore, the number of potential peptides grows rapidly, making AutoMod a powerful tool in matching peptides that are missed by conventional databank searching. To ensure that the tool is used efficiently you must take care to limit this value to a sensible figure, and to assign the peptide tolerance appropriately.

Default: By default, each peptide is allowed to contain one modification or substitution.

Specifying the likelihood of substitutions

The likelihood of each individual amino acid substitution has been calculated in the generation of the Blosum62 matrix, and is represented as a score from -4 to 11; -4 being an unlikely substitution and 11 being the most likely. For example, substitution of a methionine for a leucine has a score of 2, substitution of a tryptophan for a proline has a score of -4.

In the text box, type a value between -4 and 11. This limits the number of substitutions considered to those that have a higher value than the one specified.

Validate Results

All MSMS results can be validated. A validated peptide will contain a series of three or more consecutive y-ions.

If validation is selected, the top scoring peptide for each MSMS spectrum is returned. This could increase the requirement for manual validation of the results returned.

Selecting protein sequences for the search

Requirement: When running a one-off AutoMod analysis either protein sequences, EST sequences, or both must be specified.

If an AutoMod query is created as part of a workflow, protein sequences and EST sequences can be omitted, since the proteins and ESTs identified by any preceding databank search are used as the input for the AutoMod analysis.

Protein sequences can be typed, copied and pasted, or dragged and dropped into the text area. The sequences must be in fastA format.

Tip: fastA format sequences can be added by dragging and dropping proteins from the navigator tree or protein table in a ProteinLynx search results frame.

Selecting EST sequences for the search

Requirement: When running a one-off AutoMod analysis either protein sequences, EST sequences, or both must be specified.

If an AutoMod query is created as part of a workflow, protein sequences and EST sequences can be omitted, since the proteins and ESTs identified by any preceding databank search will be used as the input for the AutoMod analysis.

EST sequences can be typed, copied and pasted, or dragged and dropped into the text area. The sequences must be in fastA format.

Tip: fastA format sequences can be added by dragging and dropping ESTs from the navigator tree or protein table in a ProteinLynx search results frame.

De Novo Sequencing tool

De Novo sequencing enables you to determine the primary sequence of a peptide directly from its MSMS data. This is achieved by analyzing the mass differences between the peptide fragment ions. This tool facilitates the characterization of peptides whose protein or EST has not yet been entered into a databank and generates sequences that can be subsequently used in a BLAST search.

You can use the De Novo Sequencing tool to search data from any instrument that can generate fragmentation spectra: Electrospray Q-Tof, Maldi PSD and Maldi Q-Tof.

This type of analysis is primarily used as the third step in a workflow, to sequence MSMS data not matched by a Databank or AutoMod query. De Novo sequencing can also be carried out as a one-off query, where all the available fragmentation data is sequenced.

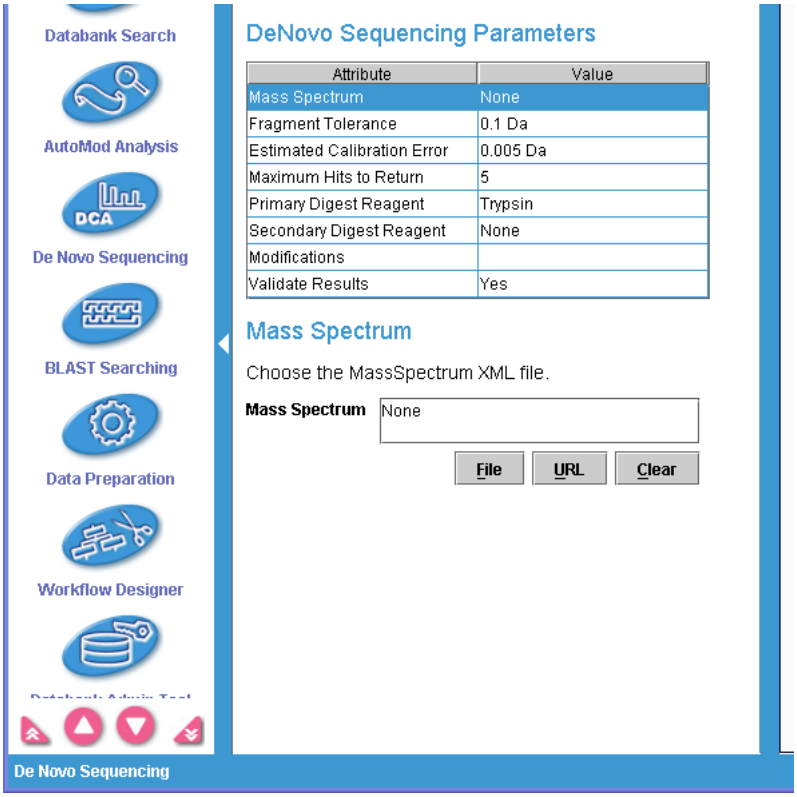
Note: Adding a De Novo query to a workflow differs only slightly from carrying out an individual search and so the following section contains information relevant to both types of experiment.

To open the De Novo Sequencing query tool, click the De Novo Sequencing icon

in the tool tray.

The De Novo Sequencing Parameters table opens in the Editor Panel of the browser.

De Novo Sequencing parameters:



To perform De Novo sequencing:

1. Click an attribute in the table (see De Novo sequencing parameters on page 14-21 for details), and then edit the value in the panel at the bottom of the table.
2. When the required fields have been edited, click the Submit button on the toolbar to start the search.

When the analysis is complete, the results are displayed in the unified results panel that is added to the desktop.

De Novo sequencing parameters

The following sections detail the attributes in the De Novo Sequencing Parameters table.

The parameters Mass Spectrum, Fragment Tolerance, Primary Digest Reagent, Secondary Digest Reagent are described in Databank search parameters on page 14-5.

Specifying the estimated calibration error

This value is fundamental to the scoring of a peptide sequence against a given fragmentation spectrum. A tight error will significantly reward well-measured data in the scoring, so it is recommended that spectra submitted are well mass measured to allow a low estimated calibration error to be set.

It is not necessary to adjust the estimated calibration error for small variations of this number in the fourth decimal place.

This value will be combined with the estimated mass measurement error for each peak. The estimated mass measurement error is calculated by the processor.

To specify an estimated calibration error, type the value into the text field, and then select the units from the combo box. Available units are Daltons (Da), and parts per million (ppm).

Specifying maximum hits to return

The Maximum Hits to Return attribute corresponds to the maximum number of De Novo sequenced peptides to return per fragmentation spectrum.

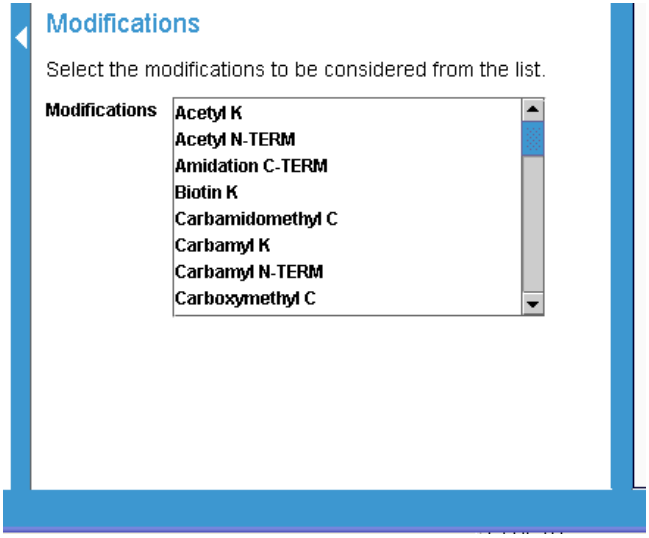
If the Validate Results feature is used, only those peptides that are validated will be returned. It is therefore possible that fewer sequences are returned for some spectra than the value specified here.

Specifying modifications to peptides

Specifying modifications is optional. By default, no modifications are applied to the peptides produced by the digest.

The Modifications list contains all the available modifier reagents.

De Novo Sequencing parameters: Modifications list:



Click a reagent in the list to specify a variable modifier that should be applied to peptides produced by the digests. To select multiple modifier reagents, use Shift+click or Ctrl+click.

Both modified and unmodified versions of each peptide will be used in the search.

Validate Results

All MSMS results can be validated. A validated peptide will contain a series of three or more consecutive y-ions.

If validation is selected, the top scoring peptide for each MSMS spectrum is returned. This could increase the requirement for manual validation of the results returned.

BLAST Searching tool

The BLAST Searching tool performs a homology search on the selected databank using the input protein/peptide sequences.

- BLAST predicts which proteins the input sequence could be a part of.
- BLAST searches can be performed as one-off searches using the BLAST search tools.
- BLAST searches can be performed using the workflow system, enabling the BLAST search to be combined with other searches.

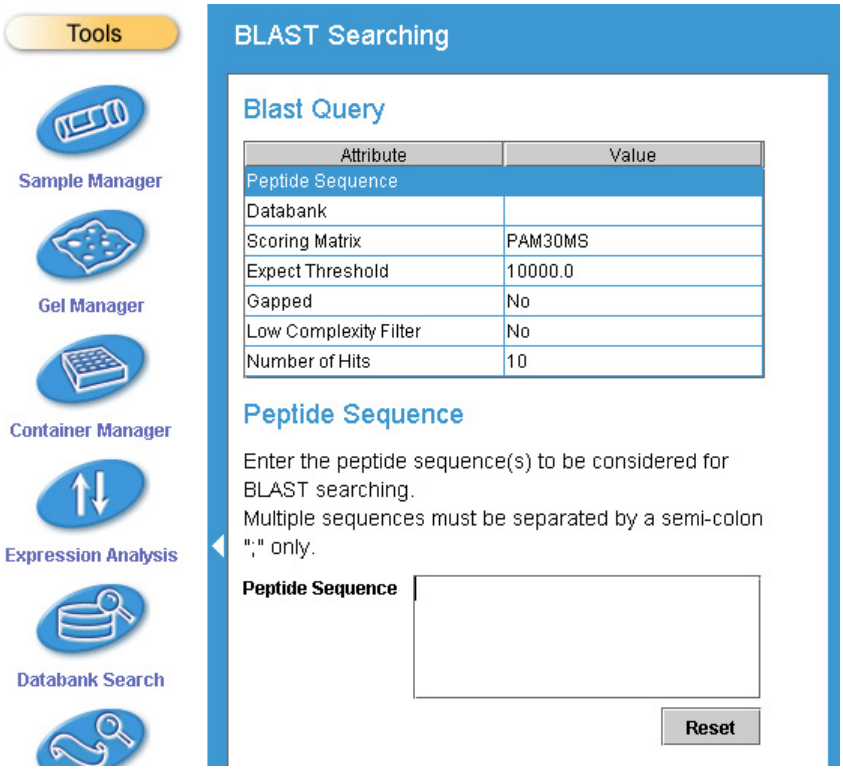
See sections on Workflow Designer (page 7-1) and Container Manager (page 5-2) for details of how to perform BLAST searches and other searches as part of an integrated workflow.

Tip: Careful use of the algorithm through automated workflows can increase coverage and confidence of the top databank search hits, while simultaneously filtering out questionable, lower-scoring hits.

To open the BLAST Searching tool, click the BLAST Searching icon in the tool tray.

The BLAST Searching Parameters table opens in the editor panel of the browser.

BLAST Searching parameters:



To perform a BLAST search:

1. Click an attribute in the table (see BLAST search parameters on page 14-24 for details), and then edit the value in the panel at the bottom of the table.
2. When the required fields have been edited, click the Submit button on the toolbar to start the search.

When the analysis is complete, the results are displayed in the BLAST results panel (see BLAST results on page 14-26).

BLAST search parameters

The following sections detail the attributes in the BLAST Searching Parameters table.

The parameter Databanks is described in Databank search parameters on page 14-5.

Peptide sequence

In the text box, type or paste one or more sequences for searching. Each sequence should be a series of amino acid identifiers, or a sequence in FASTA format, and the sequences should be separated by semicolons.

Tip: It is possible to drag and drop, or copy and paste, sequences from the results window of a search that has already been performed.

Scoring matrix

From the list, select the scoring matrix for the search.

- The PAM family of matrices were developed by Dayhoff, (see Dayhoff MO, Atlas of Protein Sequence and Structure, 5, suppl. 3 (1978)).

  PAM matrices labeled with low numbers are more suitable for looking for close relationships. PAM matrices with higher numbers are more suitable for detecting weaker similarities.

- The BLOSUM family of matrices were developed by Heinikoff and Heinikoff, (see:Henikoff S, Henikoff JG, Amino acid substitution matrices from protein blocks, Proc Natl Acad Sci USA, 89(22), 10915-9(1992))

  BLOSUM matrices with high numbers are more suitable for detecting high similarity matches. Those with lower numbers are suitable for detecting more distant relationships.

Results from De Novo searches from mass spectrometry data typically consist of short sequences, of the order of 10-30 amino acids. When BLAST searching these results it is most appropriate to use parameters which favor short, nearly exact matches.

When searching for short, nearly exact matches, a preferred matrix is PAM30. The matrix PAM30MS is based on PAM30, but with account taken for the fact that mass spectrometers cannot distinguish between certain pairs of amino acids.

Expect Threshold

Type the required expect threshold.

Each search hit returned from BLAST search has an associated “E-value”. If searching a randomly generated sequence against a database, a certain number of hits would be expected to occur simply by chance. The “E-value” of a match is an indication of how many matches of that score would be expected from that databank simply by chance. The E-value depends on the scoring matrix, the size of the databank, and the length of the query sequence.

Low expectation values are a good indication that a hit could be a true hit and has not occurred spuriously. The expect threshold is the cutoff value for the expectation values when performing a BLAST search. Setting a relatively low expectation threshold gives a stricter criterion for returned hits. Setting a high expectation threshold is more lenient with regard to hits returned.

When searching for short, nearly exact matches, a high expect threshold is appropriate.

Gapped

If the check box is selected, the BLAST search allows for gaps in the alignments in the matching process.

Low Complexity Filter

If this check box is selected, the BLAST search masks for repeats in the sequence.

De Novo analysis of mass spectrometry data typically returns results which are relatively short sequences of amino acids. Masking for repeats of such short sequences can result in very little retained data.

Number of Hits

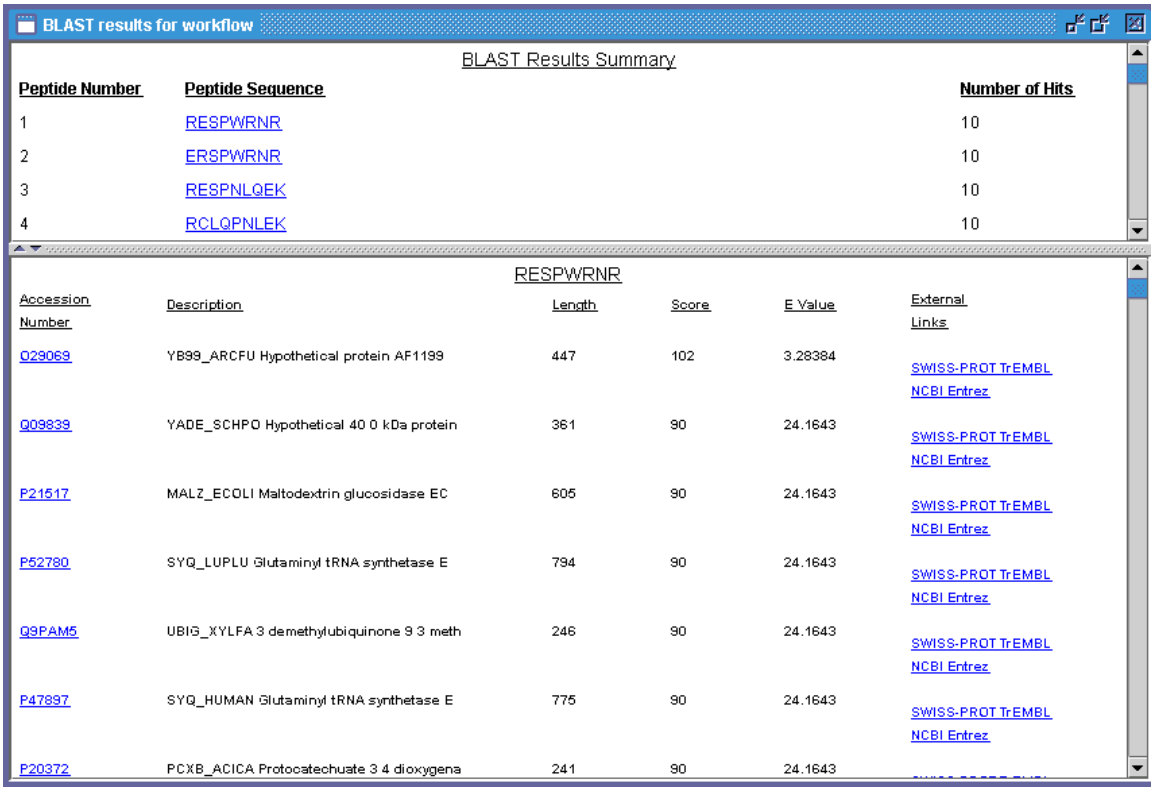
In the text box, type the maximum number of hits to be returned from the search.

BLAST results

When the search is complete, the results are returned in a BLAST results panel. The BLAST results panel is added to the results desktop which is common to this and other ProteinLynx tools.

In the example illustrated, the results panel displays the hits obtained by submitting a single sequence for BLAST searching.

BLAST results panel:


BLAST results for workflow
BLAST Results Summary
Peptide Number
Peptide Sequence
Number of Hits
1 RESPWRNR 10
2 ERSPWRNR 10
3 RESPNLQEK 10
4 RCLQPNLEK 10
RESPWRNR
Accession Number
Description
Length
Score
E Value
External Links
Q29069 YB99_ARCFU Hypothetical protein AF1199 447 102 3.28384 SWISS-PROT TrEMBL NCBI Entrez
Q09839 YADE_SCHPO Hypothetical 40 0 kDa protein 361 90 24.1643 SWISS-PROT TrEMBL NCBI Entrez
P21517 MALZ_ECOLI Maltodextrin glucosidase EC 605 90 24.1643 SWISS-PROT TrEMBL NCBI Entrez
P52780 SYQ_LUPLU Glutaminyl tRNA synthetase E 794 90 24.1643 SWISS-PROT TrEMBL NCBI Entrez
Q9PAM5 UBIG_XYLFA 3 demethylubiquinone 9 3 meth 246 90 24.1643 SWISS-PROT TrEMBL NCBI Entrez
P47897 SYQ_HUMAN Glutaminyl tRNA synthetase E 775 90 24.1643 SWISS-PROT TrEMBL NCBI Entrez
P20372 PCXB_ACICA Protocatechuate 3 4 dioxygena 241 90 24.1643

Navigating within a BLAST results panel

The BLAST results panel consists of an upper and a lower section. The upper section lists the sequences which have been BLAST searched.

Click a Peptide Sequence hyperlink in the upper section of the window. BLAST results for that sequence are displayed in the lower section of the window.

To see the alignment for a hit, scroll down in the lower section of the BLAST Results Panel. Alternatively, click on the hyperlink of one of the matches to jump to the alignment details for that hit.

15 Real Time Databank Searching

The Real Time Databank searching application allows the acquisition system, or more particularly a data-dependent acquisition (DDA), to be updated according to the results obtained from a databank search.

Specifically, if a protein is identified while a data-dependent acquisition is in progress, the software generates all the peptide masses corresponding to the identified protein. The acquisition system then uses these masses to form an exclude list to prevent any further MSMS data collection for that particular protein.

Real Time Databank searching is accessed from within MassLynx.

See also: Some familiarity with MassLynx is recommended. Refer to the MassLynx Getting Started Guide, and the MassLynx Help, for information on using the MassLynx window, sample lists, and the MassLynx queue.

You will also need to refer to the Data Acquisition sections of the MassLynx Help or relevant Operator's Guide.

Rule: Real Time Databank searching is only available for MassLynx versions 4.0 SP1 and later.

Contents:

Topic	Page
Using real time databank searching	15-2
Advanced options	15-14

Using real time databank searching

To enable real time searching of databanks there are a number of essential steps to take before the system will operate correctly.

To enable real time searching:

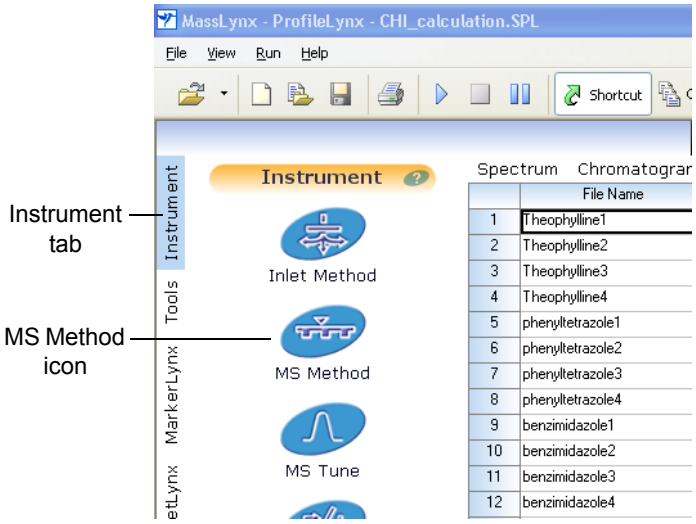
1. Ensure you have launched the Real Time Databank Searching application (Launching the Real Time Databank Searching application on page 15-2).
2. Set up the acquisition by (see Setting up a real time databank searching acquisition on page 15-8):
   - Creating a conventional MassLynx DDA acquisition method.
   - Running the ProteinLynx databank search engine microkernel.
   - Enable real time processing for processing raw data and query submission.
3. Edit raw data processing parameters according to your requirements (see Processing parameters on page 15-4).
4. Edit the databank searching parameters including setting the appropriate databank (see Searching parameters on page 15-5).
5. Start a MassLynx acquisition using the appropriate DDA method.
6. Display the databank results - Real Time Status - during the acquisition (see Real time status on page 15-7).

Launching the Real Time Databank Searching application

To launch the ProteinLynx Real Time Databank Searching application:

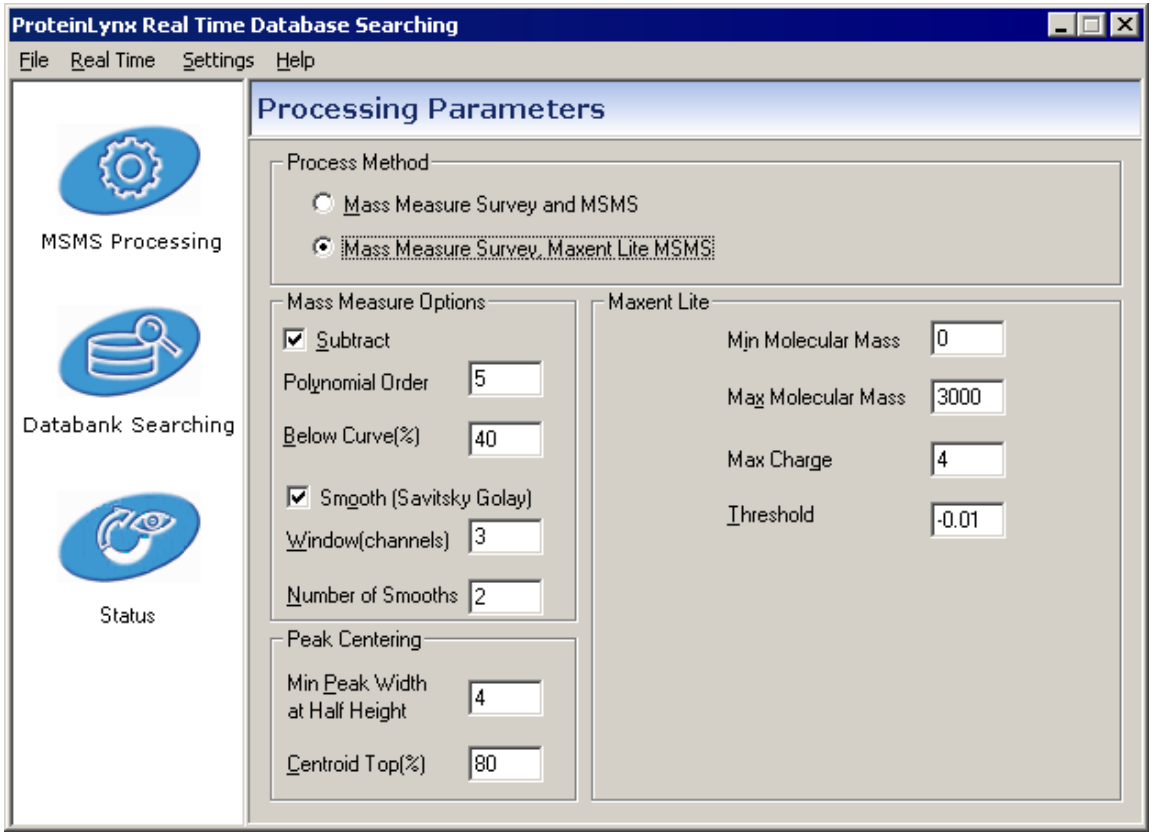
1. In MassLynx, click the Instrument tab, and then click the MS Method icon.

MS Method editor launch:



2. In the MS Method editor, click Options > ProteinLynx Real Time.

Real Time Databank Searching application:



Processing parameters

When the Real Time Databank Searching application is launched, the Processing Parameters view is usually displayed.

If this view is not displayed, click the MSMS Processing icon in the tool tray.

You can change the following parameters.

Processing parameters:

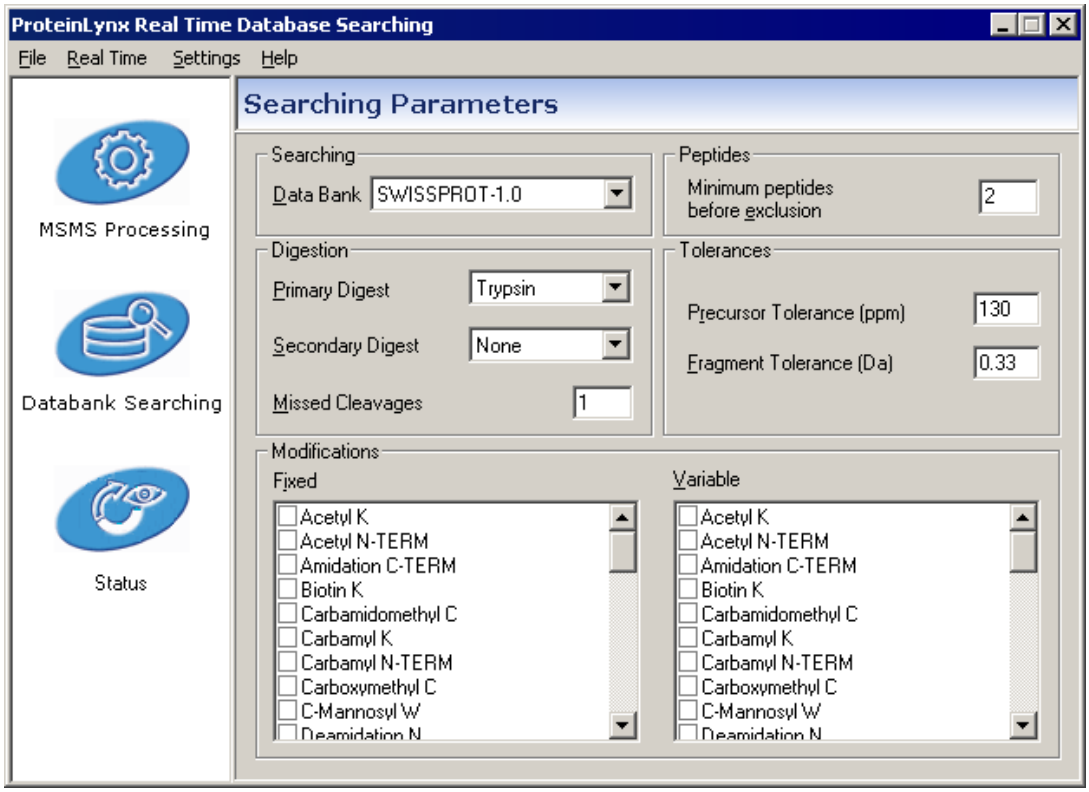
Parameter	Description
Process Method	Mass Measure Survey and MSMS – Apply the same MassLynx mass measure algorithm to both the survey scan data and the MSMS scan data. Mass Measure Survey, MaxEnt™ Lite MSMS – Apply the MassLynx mass measure algorithm to the survey data and perform MaxEnt Lite deconvolution to the MSMS data.
Subtract	Select the box to enable background subtraction of the raw data and adjust the settings according to your requirements.
Smooth	Select the box to perform Savitsky Golay smoothing of the data. Adjust the smoothing parameters according to your requirements.
Peak Centering	Adjust the parameters according to your requirements.
MaxEnt Lite	MaxEnt Lite will produce a singly charged, deisotoped spectrum for interpretation by the search engine. Type the molecular mass range of this spectrum, the maximum charge expected in the data, and a threshold setting. For the threshold setting, type a negative value for relative (percent) thresholding, or a positive value for absolute thresholding. Data below the threshold will not be considered by MaxEnt Lite.

Searching parameters

To view or edit the Searching Parameters, click the Databank Searching icon in the tool tray.

The Searching Parameters view is displayed.

Searching Parameters page:



You can change the following parameters.

Searching parameters:

Parameter	Description
Data Bank	The Data Bank drop-down list will show the available databanks. Click the one you wish to search against.
Digestion	Choose the digest reagents you wish to use when searching the data, and the number of missed cleavages.
Peptides	Type the minimum number of peptides that must match against a protein before that protein is excluded from further data acquisition.
Tolerances	Type the precursor and fragment ion tolerances to be used by the databank search engine.

Searching parameters: (Continued)

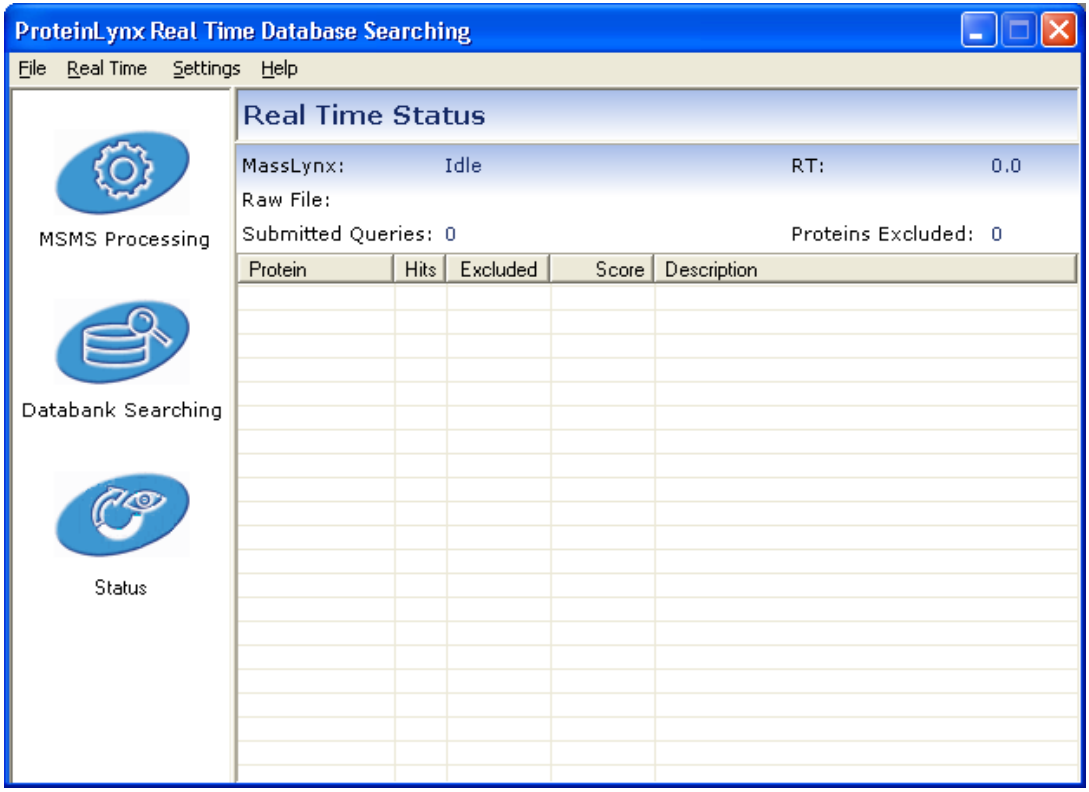
Parameter	Description
Modifications	Select and clear check boxes to set the fixed and variable modifications.

Real time status

To view the real time status, click the Status icon .

The Real Time Status view is displayed.

Real Time Status page:



The following information is displayed.

Real Time Status parameters:

Parameter	Description
MassLynx	Indicates whether MassLynx is acquiring data or idle.
RT	The retention time during an acquisition.
Raw File	The currently acquiring raw file.
Submitted Queries	The number of processed spectra that have been submitted to the search engine.
Proteins Excluded	The number of proteins that have been used to generate excluded lists.

In addition, a table of results displays and updates details of the identified proteins, including the protein name from the databank, and whether that particular protein has been excluded.

Setting up a real time databank searching acquisition

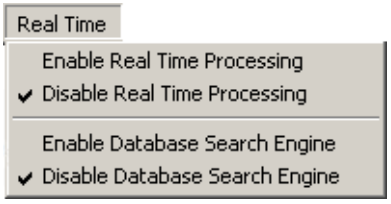
To set up a real time databank searching acquisition:

1. Create a conventional DDA acquisition from MassLynx (Setting up your DDA file on page 15-10).

   See also: If you are unsure how to do this, refer to the MassLynx Help.

2. Launch the ProteinLynx search engine: on the menu bar, click Real Time > Enable Database Search Engine. If the program is already running, there will be a tick against this menu option.

   ### Real Time menu:

   The search engine program accepts processed spectra and identifies proteins which match the spectra. If a given number of spectra (peptides, in other words) have matched to a particular protein then the

protein is ‘digested’ and an exclude mass list generated. It is possible for the user to set the number of peptides to match a protein before that protein is excluded.

Rule: These database menu items will be unavailable if you have selected remote microkernel – see Advanced options on page 15-14 for more details.

3. Click Real Time > Enable Real Time Processing. If monitoring is already enabled, there is a tick against this menu option.

   Enabling real time processing allows the system to monitor the acquisition system. If an acquisition is in progress then the raw data will be processed as it is being acquired. Each processed spectrum is then submitted to the search engine for protein identification.

4. Set the Processing and Searching Parameters (see Processing parameters on page 15-4 and Searching parameters on page 15-5), and then click File > Save to save the parameters.

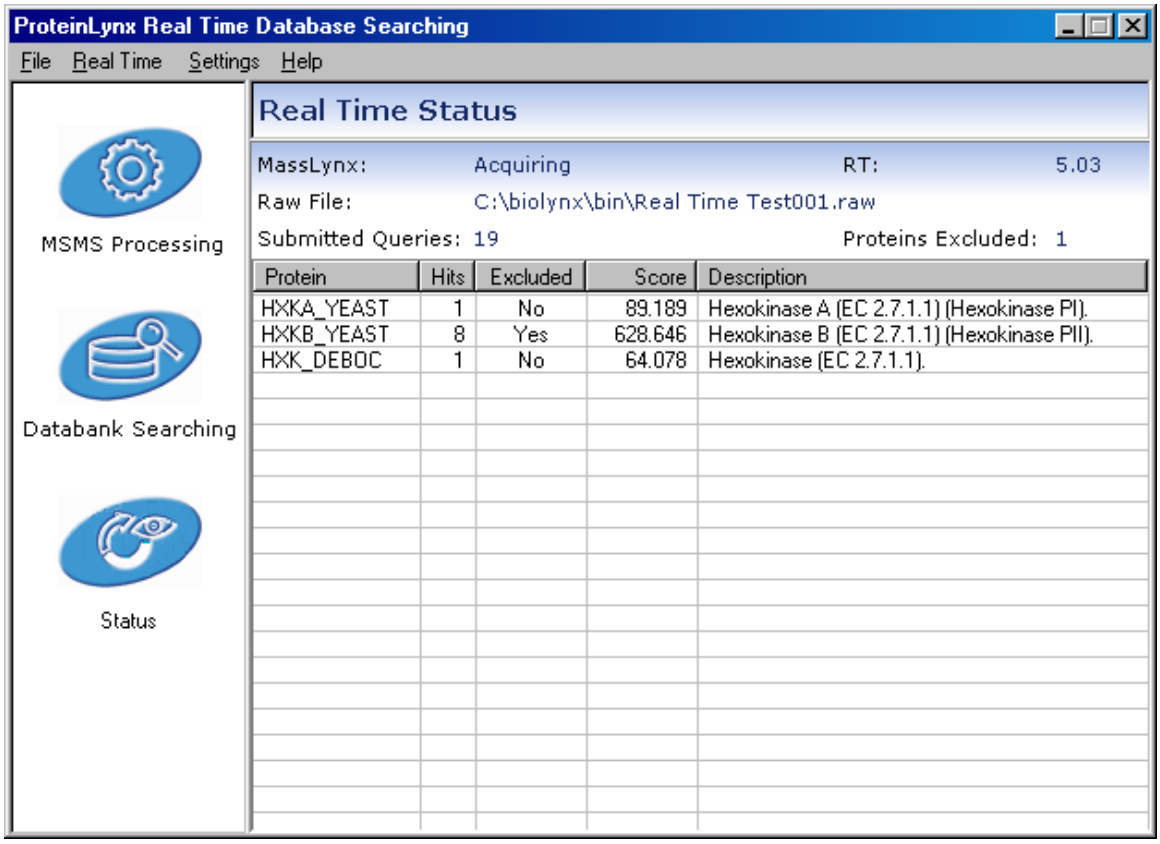
   Rule: Parameters cannot be saved if an acquisition is in progress.

5. In MassLynx, click the start button to start the acquisition.

   See also: Refer to the MassLynx Help for assistance on starting an acquisition.

6. Click the Status icon to display search results during an acquisition.

Real Time Status page with search results:


ProteinLynx Real Time Database Searching
File Real Time Settings Help
MSMS Processing
Databank Searching
Status
Real Time Status
MassLynx: Acquiring RT: 5.03
Raw File: C:\biolynx\bin\Real Time Test001.raw
Submitted Queries: 19 Proteins Excluded: 1
Protein Hits Excluded Score Description
HXKA_YEAST 1 No 89.189 Hexokinase A (EC 2.7.1.1) (Hexokinase PI).
HXKB_YEAST 8 Yes 628.646 Hexokinase B (EC 2.7.1.1) (Hexokinase PII).
HXK_DEBOC 1 No 64.078 Hexokinase (EC 2.7.1.1).

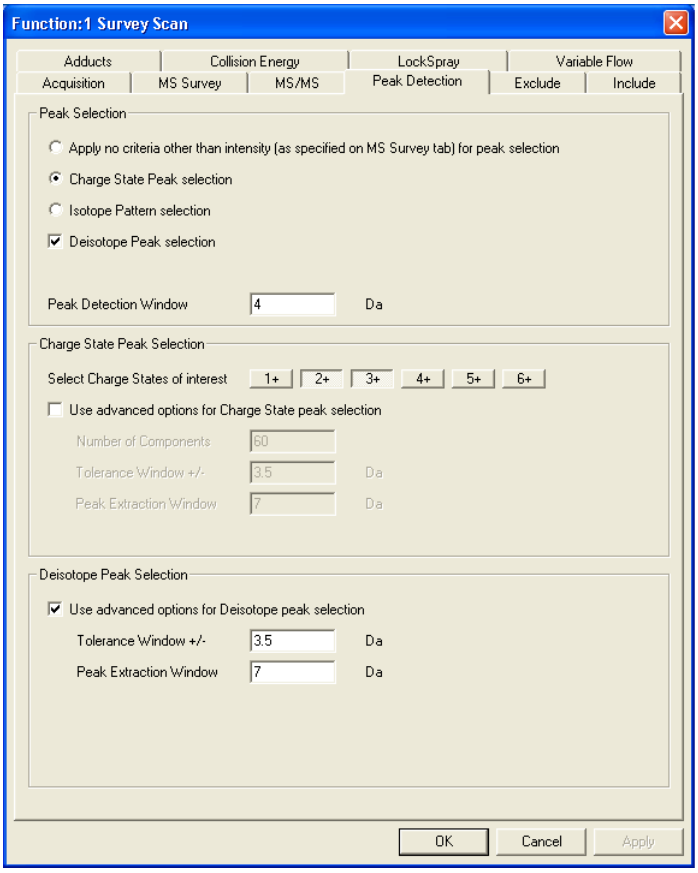
Setting up your DDA file

Real time databank searching is designed to work interactively with DDA. For this combination to work effectively the instrument needs to use de-isotope peak detection, and for this to work properly modifications to your DDA experiment need to be made.

The following graphic shows suggested settings for the Peak Detection and Exclude tabs of the DDA Survey experiment settings.

Exception: On some instruments, the settings shown below will appear in slightly different locations within the experiment dialog box. Refer to the MassLynx Help and the Operator's Guide for your instrument, using the settings below as guidelines.

Peak Detection tab:



De-isotope peak detection

For a more in depth description of the workings of Deisotope Peak Detection see the MassLynx Help.

De-isotope peak detection is enabled by selecting the Deisotope Peak selection box on the Peak Detection tab of the DDA experiment settings (Figure titled “Peak Detection tab:” on page 15-11).

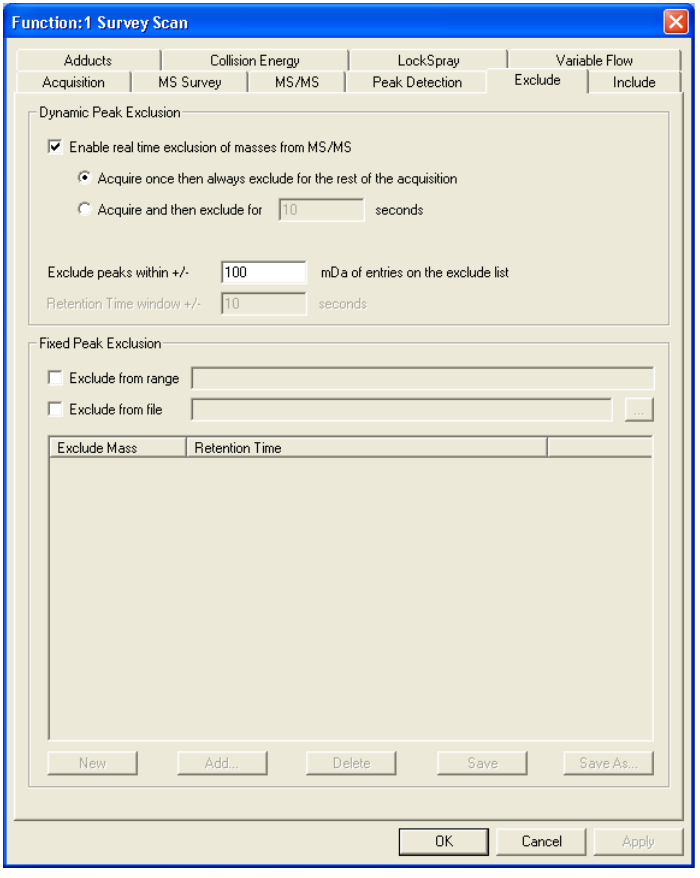
Tolerance window

The tolerance window is a window of user-defined m/z that slides up the m/z range looking for isotope clusters. Only peaks that are above the intensity threshold are considered in this routine. An ideal value for this is the distance from the tallest peak in an isotope cluster to the end of the cluster in Da (Figure titled “Peak Detection tab:” on page 15-11).

Extraction window

Once a peak has been selected by the peak detection window a section of the mass scale around the peak is taken for deisotoping. An ideal setting for this value is half the overall peak cluster size (Figure titled “Peak Detection tab:” on page 15-11).

Exclude tab:



Exclude window

The exclude window on the Exclude tab (Figure titled “Exclude tab:” on page 15-12) can then be set to 100 mDa, or lower if desired.

Other DDA experiment settings

Other settings are comparable to a normal DDA experiment.

Advanced options

The following are advanced options in the ProteinLynx Real Time Databank Searching application:

- Real time data processing
- Remote searching
- Diagnostics

Data processing

To adjust the way that the Real Time system processes data, click Settings > Real Time Processing.

You can set the following parameters.

Real Time processing setup parameters:

Parameter	Description
Start Processing After	The real time system will remain idle until the acquisition time has reached this value. **Example:** If you only expect peptides to elute after 10 minutes, set this value to 10.
Check for new peptides every	Set this time to determine how often the acquiring data is to be processed. **Example:** If this is set to 20 seconds then the raw data will be processed every 20 seconds, and if any further peptides are found they will be submitted to the microkernel search engine.

Remote searching

It is possible to process data on the acquisition PC and submit processed spectra to a search engine running on a remote PC. This can be particularly important if the acquisition PC on which MassLynx is running is of limited power.

To set remote searching:

1. Click Real Time > Disable Real Time Processing.

2. Click Settings > Microkernel Search Engine.
3. Select Microkernel Remote to enable the Microkernel URL text box.
4. Type the URL of the computer on which the microkernel search engine is running, and then click OK. You should ensure the microkernel is running on the remote PC:
   - On the remote PC, start the microkernel automatically (by starting ProteinLynx browser) or manually. See Chapter 1 - Installing ProteinLynx Global SERVER for details.
   - Run the microkernel search engine from the command by typing *PLmicrokernel.exe MassLynxURL RemoteURL*.

     Example: If the MassLynx PC has the URL 10.1.14.85 and the URL of the PC on which you are running the search engine is 10.1.11.193, type PLmicrokernel.exe 10.1.14.85 10.1.11.193.

     Requirement: You must know the URL of both this PC and the MassLynx acquisition PC.

When the program enters the wait state it is ready to take input from the MassLynx PC.

Displaying diagnostics

Diagnostic windows display processing and search information. It is not usually necessary to have these windows visible.

To display the diagnostic windows, click Help > Show Diagnostics.

Caution: Do not close the diagnostic windows by clicking the close buttons at the top right corner, as doing so can cause the applications to terminate. Instead, click Help > Hide Diagnostics.

If you have a local microkernel search engine, three diagnostic windows are displayed:

- PLmicrokernel search window – for displaying the state of the database search engine.
- process_kernel window – for displaying the state of the raw data processing module.
- rtdb_monitor window – for displaying the state of the module responsible for monitoring processed spectra and submitting these spectra to the microkernel.

Rule: These windows will only be displayed if you have enabled the search engine and enabled real time data processing.

16 Using MS^E for qualitative proteomics

If a Q-Tof Premier instrument is used, MS^E data can be acquired. This data can then be used in a protein identification experiment.

See also: MS^E data can be analyzed in Expression Analyses, configured in PLGS. If the optional Waters Protein Expression System is being used, analyses can also be configured for MS^E data acquired from samples without isotope labels. See the Waters Protein Expression System Operator's Guide for more details.

Contents:

Topic	Page
What is MS^E?	16-2
Creating an MS^E method file	16-3
Running an MS^E experiment	16-7

What is MS^E?

If a Q-Tof Premier instrument is being used, MS^E data can be acquired. When acquiring MS^E data, two MS functions are used in an alternating fashion:

- MS - one function is acquired in Tof-MS mode at a low collision energy (typically 4 eV) during which no fragmentation occurs to the precursor ions.
- MS^E- a second function is acquired, also in Tof-MS mode, during which the collision energy is linearly ramped between two user-defined energies (typically 15 eV to 40 eV). This induces fragmentation of any species present in the gas cell at that time.

Therefore, during the time course of the experiment, the Q-Tof Premier acquires data at low energy before stepping to an elevated collision energy, where it performs a collision energy ramp. Also, at a user-defined time, a reference scan is sampled from the NanoLockSpray reference sprayer.

Creating an MS^E method file

The low and elevated collision energies are set up from within the MS Method editor in MassLynx.

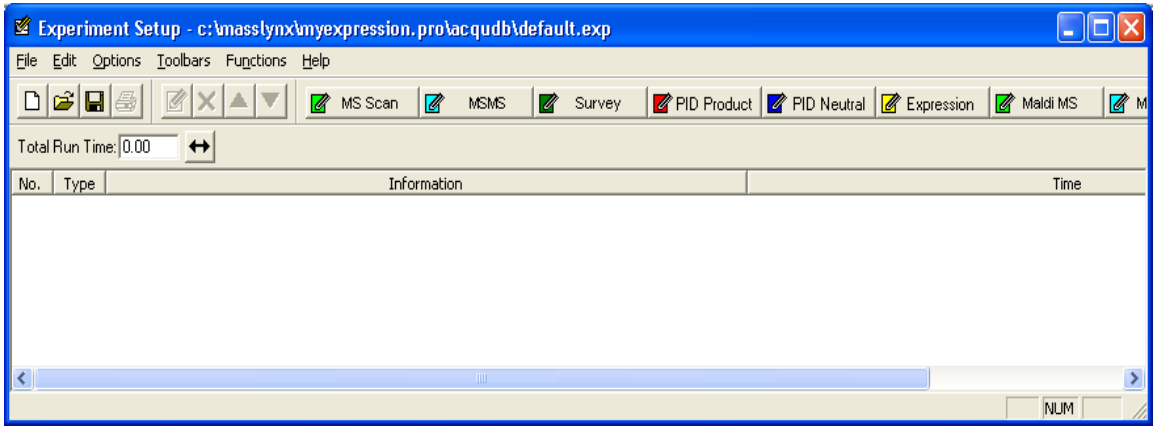
The ideal values to set for an experiment can vary depending on your hardware setup. The values shown in the screen shots that follow are suggested when using Atlantis 75µm or 300µm columns with a nanoACQUITY UPLC. Suggested values when using a BEH 75µm column are also given.

In all circumstances, some experimentation might be necessary to find the optimal values for your requirements.

To create an MS^E experiment file:

1. In the MassLynx shortcut bar, click MS Method.

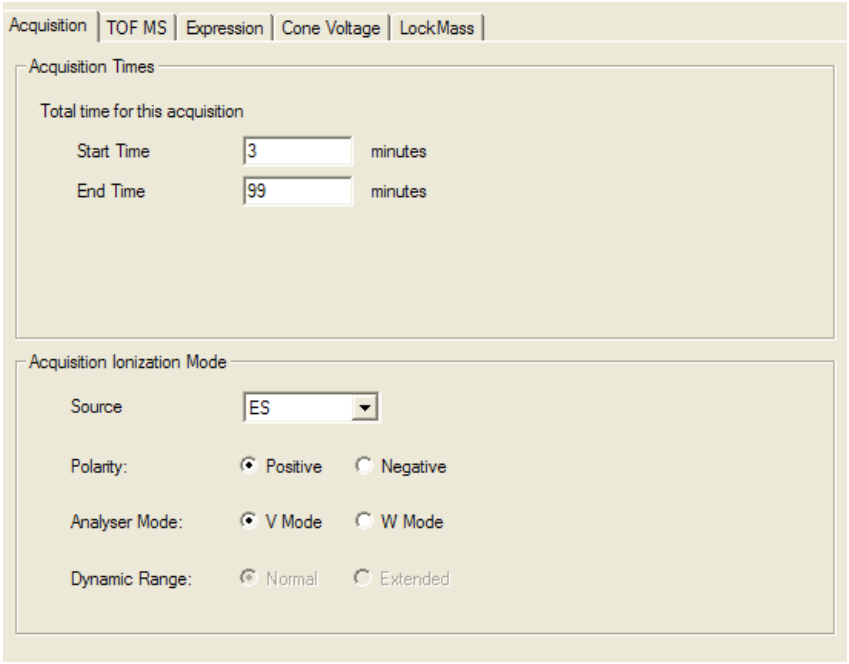
MS Method editor:



2. Delete the default function that is present in the function list.
3. Click Expression to open the Expression function editor.
4. On the Acquisition tab enter the values as shown.

   Tip: The Start and End times mirror the LC gradient. The times shown below relate to a 90 minute gradient.

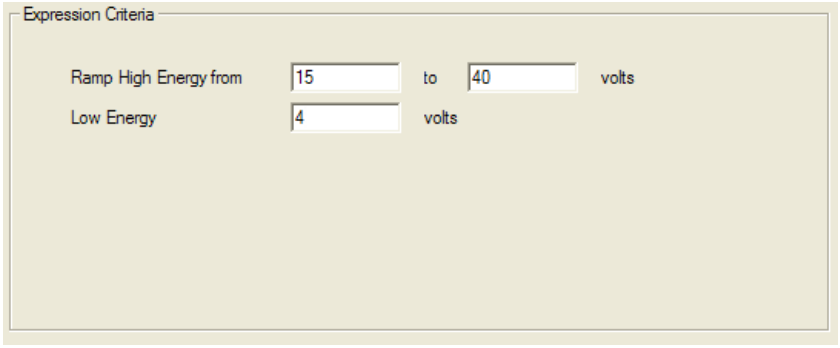
Acquisition tab:



Recommendation: When using BEH 75µm columns, a start time of 10 minutes and an end time of 75 minutes is suggested for a 60 minute LC gradient.

5. Click the Expression tab.
6. Enter the low collision energy value and the ramp for the elevated collision energy.

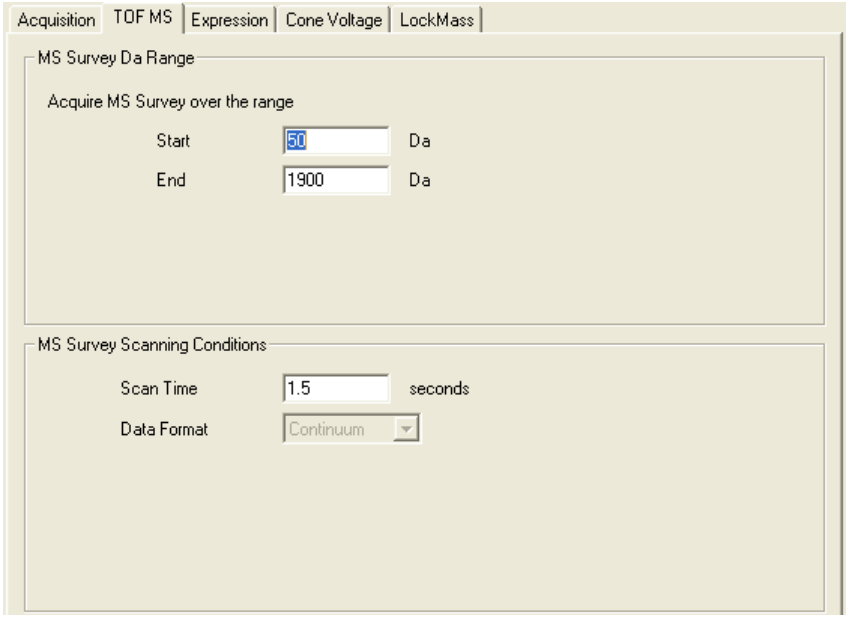
Expression tab:



The ramp for High Energy is typically set to 15 eV to 40 eV.

7. Click the TOF MS tab and enter the values as shown below.

TOF MS tab:



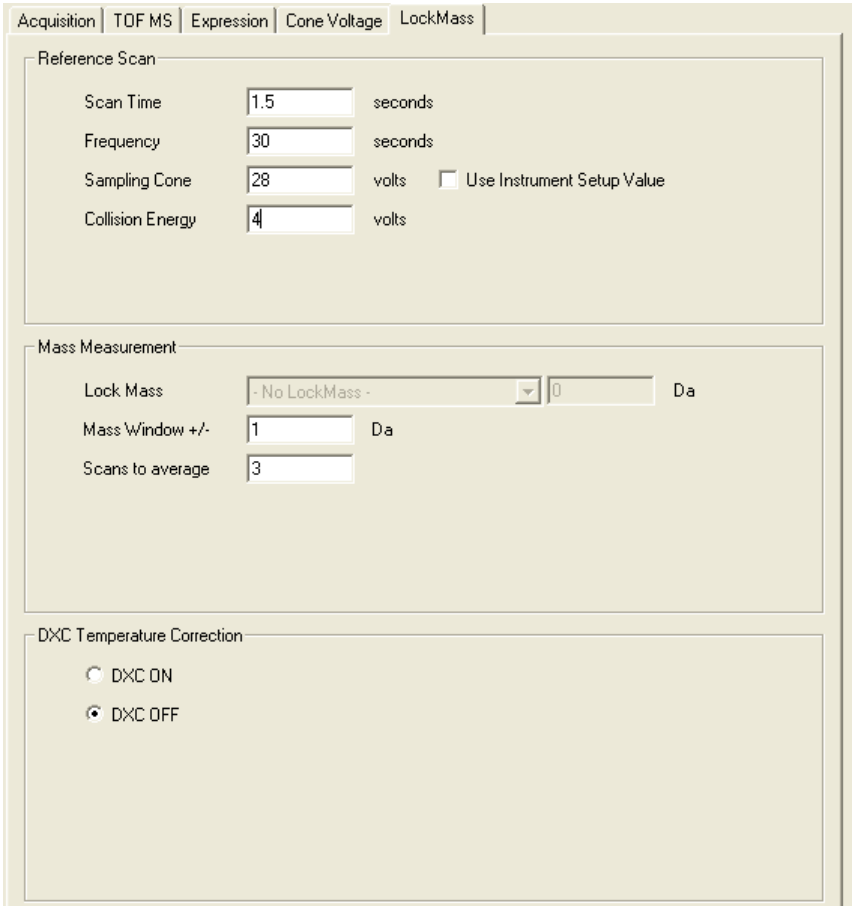
Tip: The mass range over which you wish to acquire data is typically 50 m/z to 1990 m/z.

Recommendation: When using BEH 75µm columns, a scan time of 0.6 seconds is suggested.

8. Click the LockMass tab, and then enter the values as shown below.

   Rule: The Reference Scan section of this tab is available only if the Tune window > Mode > LockSpray option is checked. Mass accuracy, and therefore Lock Spray, is an integral part of the Expression approach to data acquisition.

   LockMass tab:



   Recommendation: When using BEH 75µm columns, a scan time of 0.6 seconds is suggested.

9. Click OK.

10. In the method editor click File > Save As, and then save the experiment file with an appropriate name.

Running an MS^E experiment

All experiments are carried out through the MassLynx sample list.

See also: For information on configuring and using the sample list, refer to the MassLynx Help.

Necessary sample list fields

Only six columns are required within the sample list to carry out an MS^E acquisition:

- File Name (FILE_NAME) – each raw data file must have a file name.
- File Text (FILE_TEXT) – describes what the sample is.
- MS File (MS_FILE) – the MS^E/Expression method file.
- Inlet File (INLET_FILE) – the method file for nanoACQUITY.
- Bottle (SAMPLE_LOCATION) – position in autosampler to take sample.
- Inject Volume (INJ_VOL) – amount to inject.

Tip: As column names are configurable, they could differ from those given above. The field IDs (given in brackets above) will remain the same whatever the name of the column.

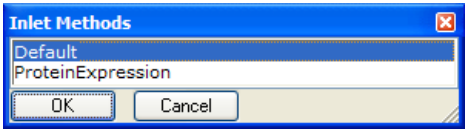
To add a method file:

1. Double-click in the MS File cell to open the Select File dialog box.
2. Choose a previously saved file MS^E method file, such as that created in the previous section, Creating an MS^E method file.
3. Click OK.

Result: The MS file is added to the sample list.

To add an inlet file:

1. Double-click in the Inlet File cell to open the Inlet Methods dialog box.



2. Click a previously saved inlet method file.

3. Click OK.

Result: The inlet method file is added to the sample list.

To run the sample list:

1. Click to start the acquisition.
2. In the Start Sample List Run dialog box, select Acquire Sample Data.
3. In the Samples frame, specify the samples to run.
4. Click OK.

When the acquisition has finished the raw data can be processed in ProteinLynx Global Server.

A Quick Start Tutorials

The following sections cover several common tasks that you might perform using PLGS. It is recommended that you are familiar with the software before attempting these procedures. Refer to Chapter 5 – Specifying samples, vials, and plates with Container Manager and all other chapters for details of how to use the software.

Ensure that PLGS is running on the computer you are using, and also on the server if one is being used. For information on how to start PLGS, see Chapter 5 – Installing ProteinLynx Global SERVER.

Contents:

Topic	Page
Creating a project and processing acquired data files	A-2
MALDI test procedure	A-5
Acquiring Q-Tof MSMS data	A-14
Adding a new databank	A-25

Creating a project and processing acquired data files

For further information see Chapter 5 – Specifying samples, vials, and plates with Container Manager.

Setting samples

To set samples:

1. Click Sample Manager.

   Note: Sample in this context refers to a batch or bottle of analyte, as distinct from a single RAW file, or line on a MassLynx sample list.

2. Click File > New Project.
3. Type a project name, and then click OK.
4. In the navigator tree, click Original Samples, and then right-click.
5. Click Add New Sample.
6. Click No to the question 'Add new sample to vial'?

   Rule: For MALDI the Target Plate container type is used instead.

7. Annotate the relevant fields with any required sample information.

   To input information, click the required field, and then type in the text box.

   Tip: The text box is active even if no flashing cursor is visible.

Setting the target plate

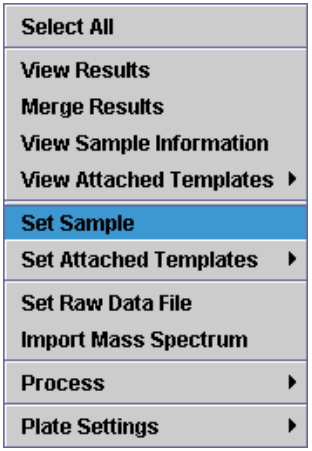
To set the target plate:

1. Click Container Manager.
2. Click Target Plates, and then right-click.
3. Click New Target Plate.
4. Type a title for the plate.

   Requirement: For MALDI HT this should match the barcode on the plate to be analyzed.

5. In the navigator tree, expand the Target Plate node, and then click the plate you created.
6. Drag across the target plate to highlight the spots corresponding to your data files.
7. Right-click anywhere in the target plate.

Target Plate pop-up menu:



8. Click Set Sample to associate the spots with the sample record previously created.
9. Select some or all of the spots again, right-click, and then click Set Raw Data File.
10. In the Select File dialog box, choose the data files to be processed, and then click OK.
11. Select some or all of the spots again, and then right-click.
12. Click Set Attached Templates > Processing Parameters.
13. Click Choose new Processing Parameters Template from file, and then choose the parameter file from disk.

    Requirement: To create and alter processing parameters, the Data Preparation tool must be used (see Getting started with the Data Preparation tool on page 8-2).
14. Select some or all of the spots again, right-click, the click Set Attached Templates > Workflow Template to RAW data.

15. Click Choose new Workflow Template from file, and then choose the workflow template from disk.

    Requirement: To create and alter workflow parameters the Workflow Designer tool (Creating a workflow template on page 7-5) must be used.

    The system is now ready to process and search.

16. Select the spots again, right-click, and then click Process > Latest RAW data.

Results:

- Progress is indicated on the status bar.
- The interface will be updated as results are returned from the server.

You can refresh the view periodically by clicking File > Update.

MALDI test procedure

Spot 24 wells of ADH with ACTH lockmass as per the installation specification.

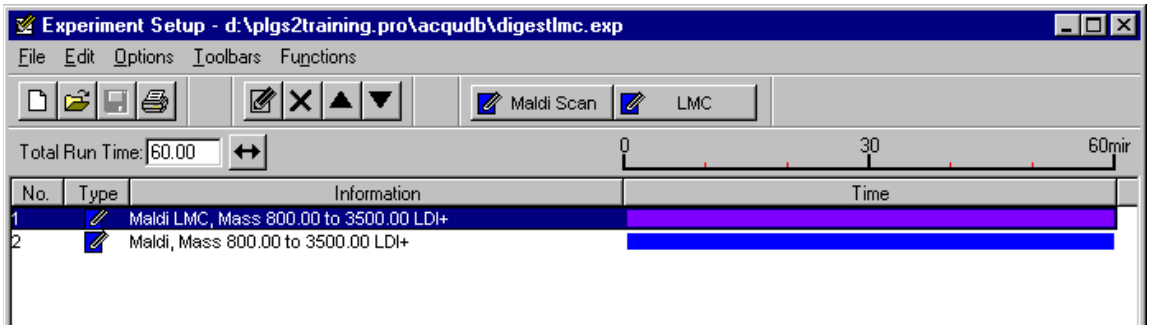
For further information see Chapter 5 – Specifying samples, vials, and plates with Container Manager.

Setting the target plate

To set the target plate:

1. Create a new MassLynx project as described in the MassLynx Help.
2. Create an MS Method File.

MS Method parameters:



3. Create a new PLGS project (see Importing and viewing PLGS sample lists on page 5-3). Enter the name of the project as PLGS2Training.
4. Click Container Manager and create a new target plate as described in Creating a new vial, microtitre or target plate on page 5-9.
5. Name the target plate.

   Tip: If using MALDI HT, use the barcode on the plate.
6. A new target plate is displayed. Drag over the spots that contain the sample.
7. Right-click on the selected wells, and then click Set Sample (see Setting a sample on page 5-11).
8. Click OK. The wells change color.

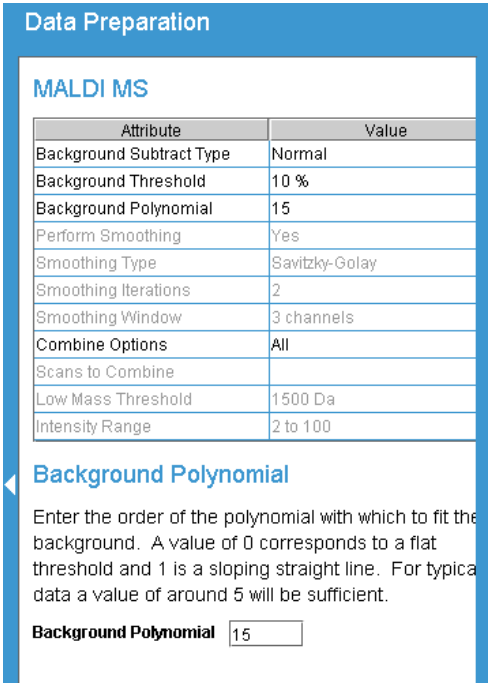
Setting processing parameters

To set processing parameters:

1. Click Data Preparation.
2. Click File > New.
3. Select Maldi MS, and then click .

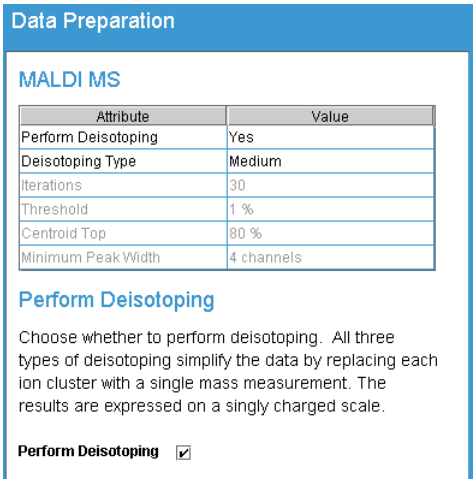
   Result: A new Processing Parameters template is opened (see MALDI Q-Tof MSMS on page 8-5).
4. Name the Processing Parameters template MALDIPP.
5. In the Mass Accuracy attributes, set the Calibration Type to External.
6. Set the External Lock Mass as 2465.1989 Da (ACTH).
7. Enter values for the Noise Reduction attributes, as shown below.

Noise Reduction attributes:



8. Enter values for the Deisotoping and Centroiding attributes, as shown below.

Deisotoping and Centroiding attributes:



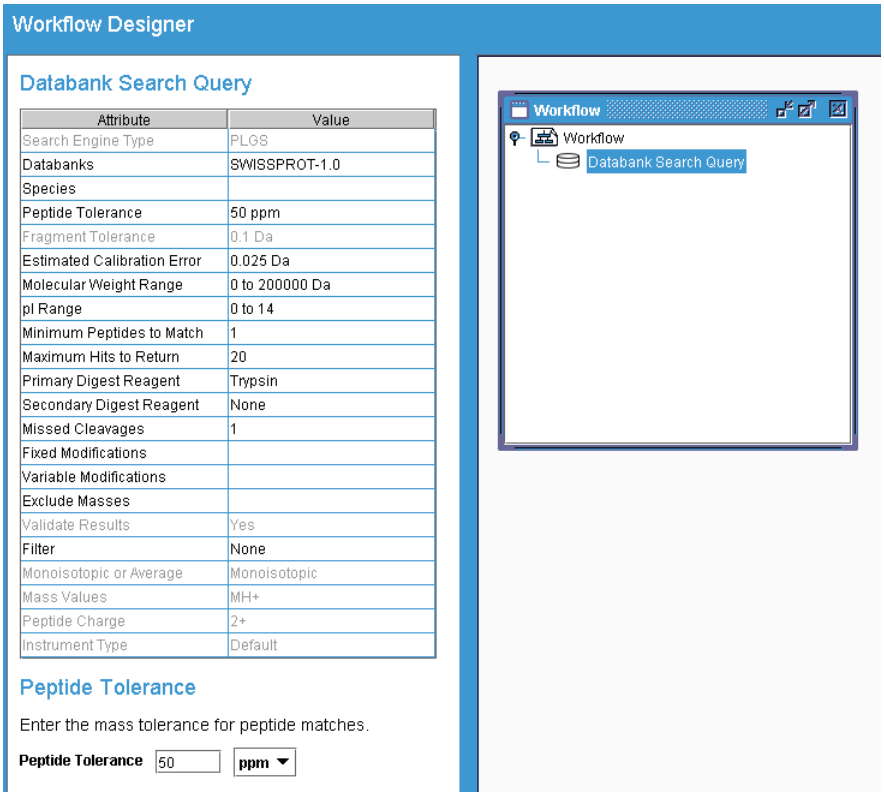
9. Click File > Save As.
10. In the Save As dialog box, save with the file name MALDIPP.

Creating a workflow

To create a workflow:

1. Click Workflow Designer (see Chapter 7 – Defining templates for searching with Workflow Designer).
2. Click File > New.
3. Select PMF, and then click .
4. Right-click the Workflow node, and then click Add > Databank Search.
5. Set the Databank Search Query parameters, as shown below.

Databank Search Query parameters:



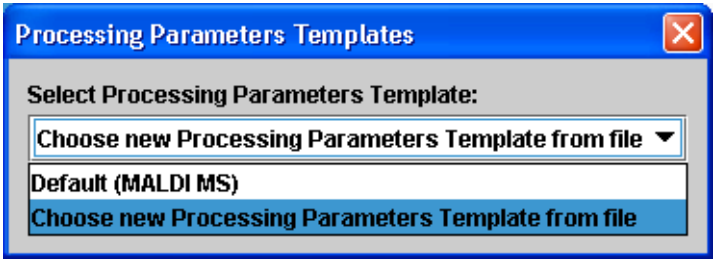
6. Select File > Save As. Name the workflow MALDIWF.

Attaching the data processing parameters

To attach the data processing parameters:

1. In Container Manager, expand the navigator tree so that the Default (MALDI MS) node, directly below the target plate name, is displayed (see Adding processing parameters templates on page 5-21).
2. Click, and then right-click, the Default (MALDI MS) node.
3. Click Change Processing Parameters.

Processing Parameters Templates dialog box:


Processing Parameters Templates
Select Processing Parameters Template:
Choose new Processing Parameters Template from file
Default (MALDI MS)
Choose new Processing Parameters Template from file

4. Click Choose new processing parameters template from file, and then click OK.
5. Click the processing parameter file, MALDIPP.xml, that you created earlier (see Setting processing parameters on page A-6), and then click Open.

Attaching the workflow file

To attach the workflow file:

1. In Container Manager, highlight all the wells on the plate for which you have set samples (see Setting the target plate on page A-5) by dragging a rectangle over them. Right-click.
2. Select Set Attached Templates > Workflow Template to Mass Spectrum.
3. Click OK, to Choose a new Workflow Template from file.
4. Click the MALDI workflow file, MALDIWF.xml, that you created earlier (see Creating a workflow on page A-7), and then click Open.

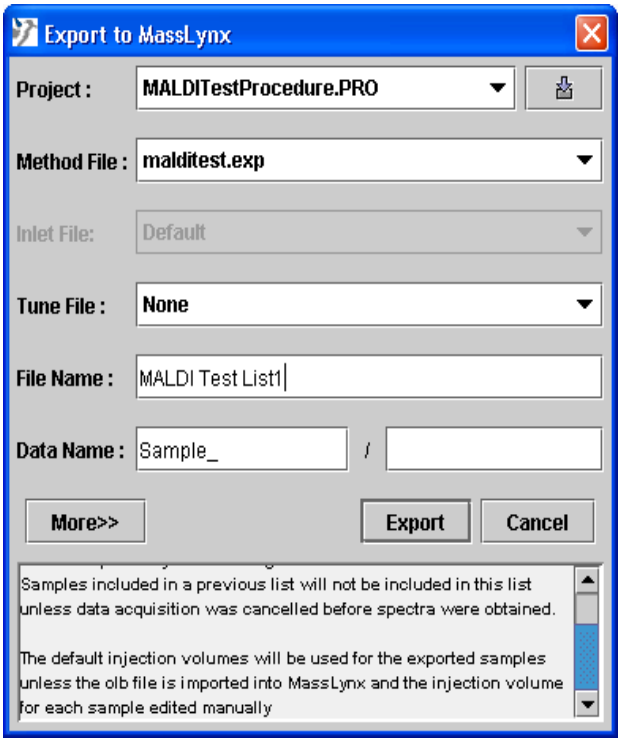
Exporting the sample list to MassLynx

For further details see Exporting a sample list to MassLynx on page 5-29.

To export the sample list:

1. In Container Manager, right-click on the target plate node, and then click Export Sample List to MassLynx.

Export to MassLynx dialog box:


Export to MassLynx
Project :
MALDITestProcedure.PRO
Method File :
malditest.exp
Inlet File:
Default
Tune File :
None
File Name :
MALDI Test List1
Data Name :
Sample_
/
More>>
Export
Cancel
Samples included in a previous list will not be included in this list unless data acquisition was cancelled before spectra were obtained.
The default injection volumes will be used for the exported samples unless the olb file is imported into MassLynx and the injection volume for each sample edited manually

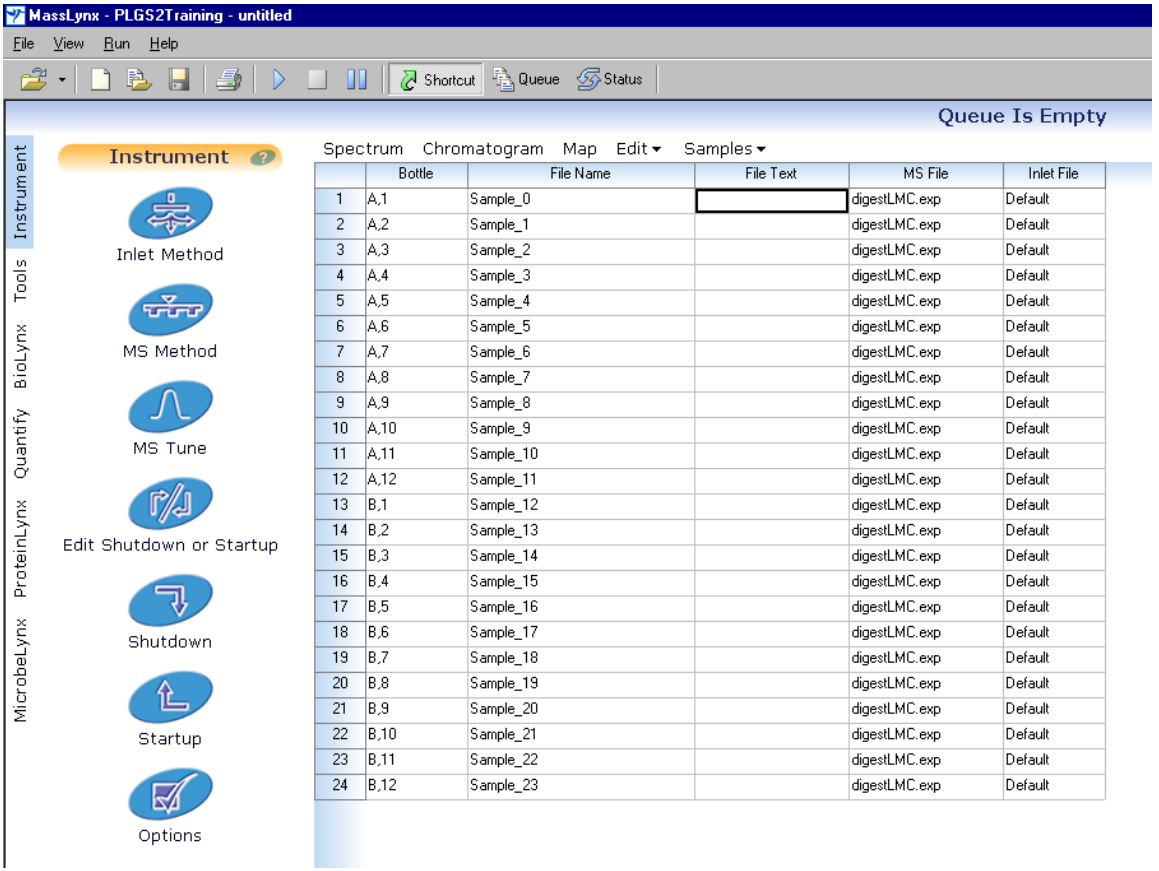
2. Specify the MassLynx project from which the data is to be acquired.
3. If more than one MS Method is stored in the MassLynx project, use the drop-down list to specify the correct file.

   Tips:

   - The File name can be the same as the target plate name.
   - The MS Data name can be changed to any text, such as digest_0, adh_0.
4. Click Export.
5. In MassLynx click File > Import Worksheet.
6. The file created by PLGS is stored in the MassLynx project. Browse to the file, and then click Open.

Result: The MassLynx sample list is updated with the information from PLGS. Data can now be acquired in the normal way.

Example MassLynx sample list:



	Bottle	File Name	File Text	MS File	Inlet File
1	A,1	Sample_0		digestLMC.exp	Default
2	A,2	Sample_1		digestLMC.exp	Default
3	A,3	Sample_2		digestLMC.exp	Default
4	A,4	Sample_3		digestLMC.exp	Default
5	A,5	Sample_4		digestLMC.exp	Default
6	A,6	Sample_5		digestLMC.exp	Default
7	A,7	Sample_6		digestLMC.exp	Default
8	A,8	Sample_7		digestLMC.exp	Default
9	A,9	Sample_8		digestLMC.exp	Default
10	A,10	Sample_9		digestLMC.exp	Default
11	A,11	Sample_10		digestLMC.exp	Default
12	A,12	Sample_11		digestLMC.exp	Default
13	B,1	Sample_12		digestLMC.exp	Default
14	B,2	Sample_13		digestLMC.exp	Default
15	B,3	Sample_14		digestLMC.exp	Default
16	B,4	Sample_15		digestLMC.exp	Default
17	B,5	Sample_16		digestLMC.exp	Default
18	B,6	Sample_17		digestLMC.exp	Default
19	B,7	Sample_18		digestLMC.exp	Default
20	B,8	Sample_19		digestLMC.exp	Default
21	B,9	Sample_20		digestLMC.exp	Default
22	B,10	Sample_21		digestLMC.exp	Default
23	B,11	Sample_22		digestLMC.exp	Default
24	B,12	Sample_23		digestLMC.exp	Default

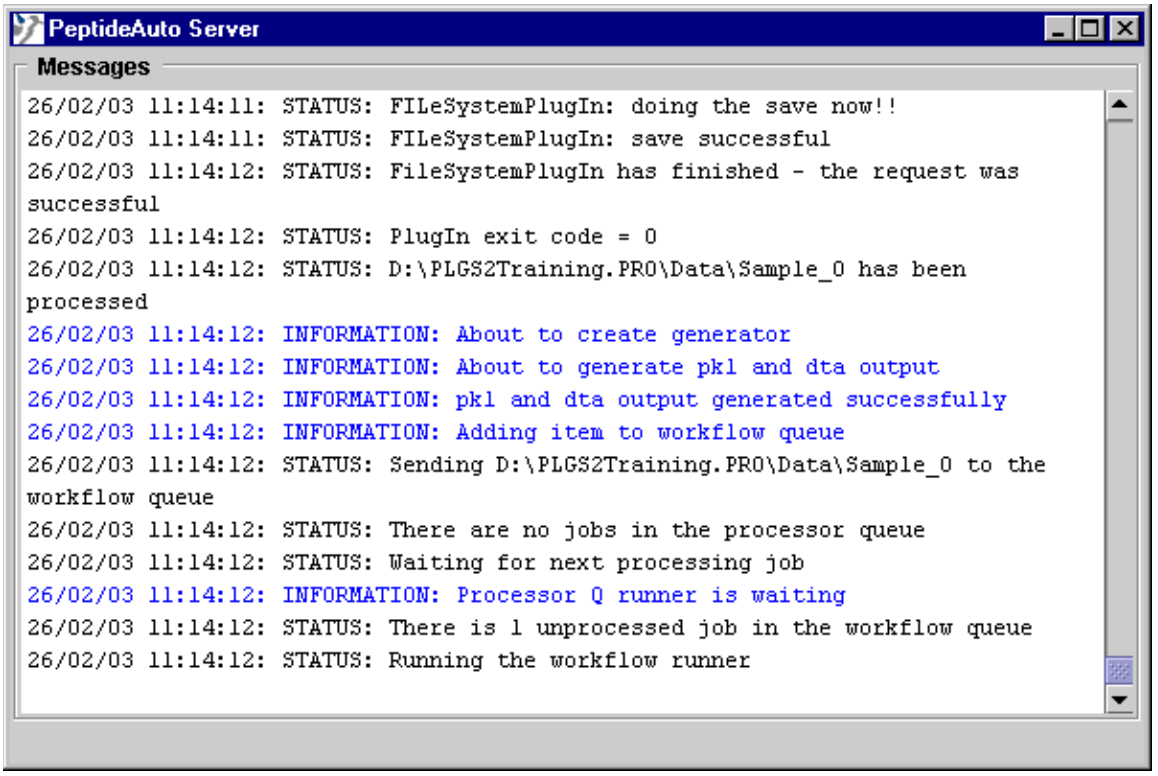
Acquiring data

To acquire data:

1. In the main MassLynx window, click to open the Start Sample List Run dialog box.
2. Select Acquire Sample Data and Auto Process Samples.
3. Click OK.
4. The PeptideAuto Server dialog box opens, which monitors the progress of the acquisition. MassLynx starts to acquire and process data.

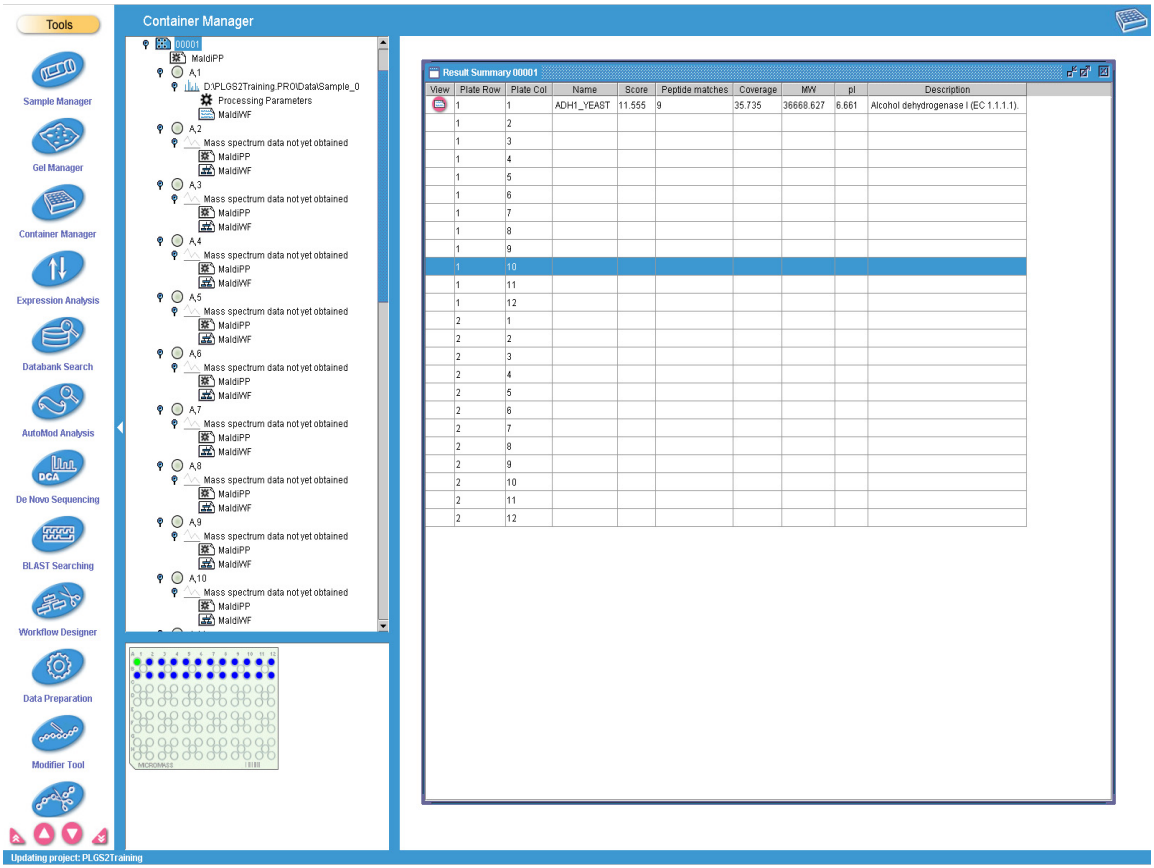
Tip: The search engine that is active in PLGS when the PeptideAuto window is opened will be the search engine used. If you wish to change the search engine, close PeptideAuto, change the search engine in PLGS, and then open PeptideAuto again.

PeptideAuto Server display:


PeptideAuto Server
Messages
26/02/03 11:14:11: STATUS: FILeSystemPlugIn: doing the save now!!
26/02/03 11:14:11: STATUS: FILeSystemPlugIn: save successful
26/02/03 11:14:12: STATUS: FileSystemPlugIn has finished - the request was successful
26/02/03 11:14:12: STATUS: PlugIn exit code = 0
26/02/03 11:14:12: STATUS: D:\PLGS2Training.PRO\Data\Sample_0 has been processed
26/02/03 11:14:12: INFORMATION: About to create generator
26/02/03 11:14:12: INFORMATION: About to generate pkl and dta output
26/02/03 11:14:12: INFORMATION: pkl and dta output generated successfully
26/02/03 11:14:12: INFORMATION: Adding item to workflow queue
26/02/03 11:14:12: STATUS: Sending D:\PLGS2Training.PRO\Data\Sample_0 to the workflow queue
26/02/03 11:14:12: STATUS: There are no jobs in the processor queue
26/02/03 11:14:12: STATUS: Waiting for next processing job
26/02/03 11:14:12: INFORMATION: Processor Q runner is waiting
26/02/03 11:14:12: STATUS: There is 1 unprocessed job in the workflow queue
26/02/03 11:14:12: STATUS: Running the workflow runner

5. To display results in PLGS, click the target plate node. The results browser opens.
6. As the data is acquired, the results in PLGS can be periodically updated, by one of the two following methods:
   - Click File > Update, or
   - Click on the toolbar.

PLGS with partially acquired sample list:



For further details on viewing results see Chapter 6 – Viewing results in the Results Browser.

Acquiring Q-Tof MSMS data

In this example one sample of hemoglobin digest is used, with glu-fibrinopeptide B (GFP) and erythromycin, infused by means of LockSpray, used as lock mass.

Setting the microtitre plate

To set the microtitre plate:

1. Create a new MassLynx project as described in the MassLynx Help.
2. Create an MS Method file and LC gradient files in the MassLynx project.
3. Create a new PLGS project (see Importing and viewing PLGS sample lists on page 5-3). Set the name of the project as Q-Tof MSMS.
4. Click Container Manager and create a new microtitre plate as described in Creating a new vial, microtitre or target plate on page 5-9. Name the microtitre plate Q-Tof MSMS.
5. Click the plate you have created, and then drag over the spot that contains the sample.
6. Right-click the selected well, and then click Set Sample (see Setting a sample on page 5-11).
7. Click OK. The well changes color.

Setting processing parameters

To set the processing parameters:

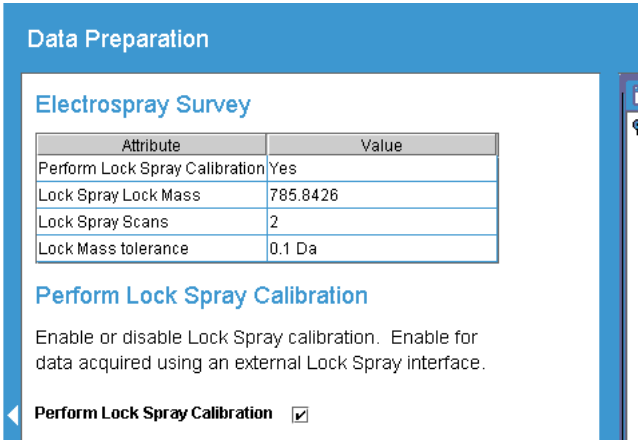
1. Click Data Preparation.
2. Click File > New.
3. Select Electrospray DDA, and then click .
4. Give the Processing Parameters the title "Data prep <*current date*>".

Each attribute set (Mass Accuracy, Noise Reduction, Deisotoping and Centroiding) has two attribute panels: Electrospray Survey and MSMS.

5. In the Mass Accuracy – Electrospray Survey panel, set the attribute Perform Lock Spray Calibration to Yes.

   Rule: The Lock Spray Lock Mass of 785.8426 Da/e – the doubly charged ion of GFP – is default in the software.

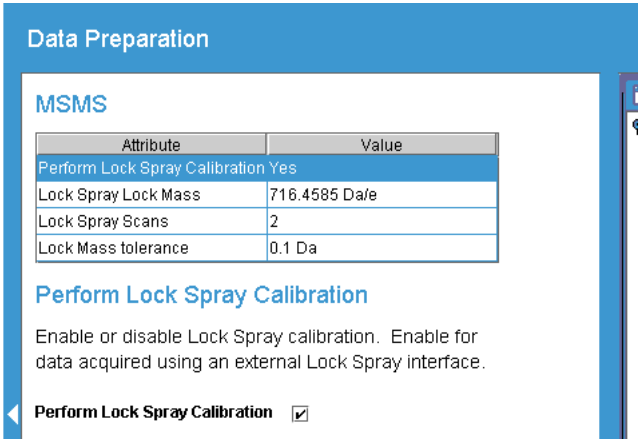
Mass Accuracy attributes – Electrospray Survey lock spray:



6. In the Mass Accuracy – MSMS panel, set the attribute Perform Lock Spray Calibration to Yes.

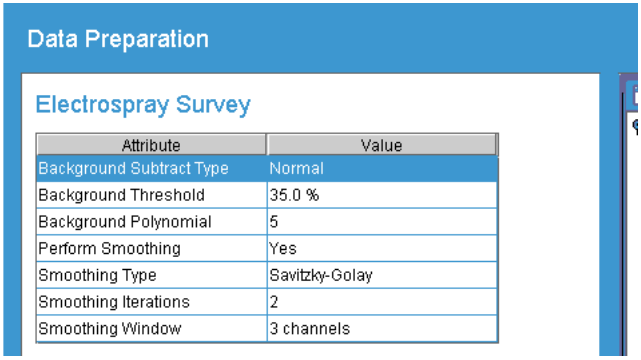
   Tip: The Lock Spray Lock Mass of 716.4585 Da/e – the single charged ion of erythromycin – is the default.

Mass Accuracy attributes – MSMS lock spray:



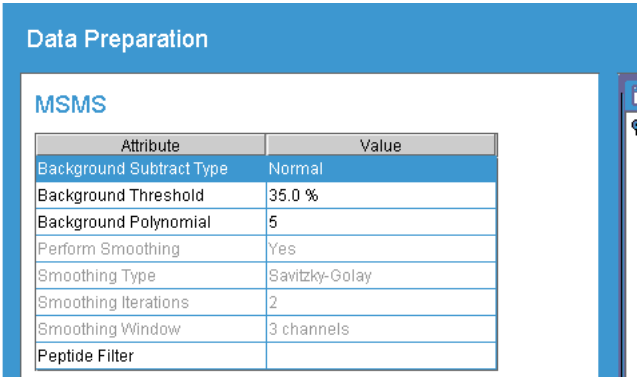
7. Set the Noise Reduction attributes in the Electrospray Survey and MSMS panels, as shown below.

Noise Reduction attributes – Electrospray Survey panel:



Attribute	Value
Background Subtract Type	Normal
Background Threshold	35.0 %
Background Polynomial	5
Perform Smoothing	Yes
Smoothing Type	Savitzky-Golay
Smoothing Iterations	2
Smoothing Window	3 channels

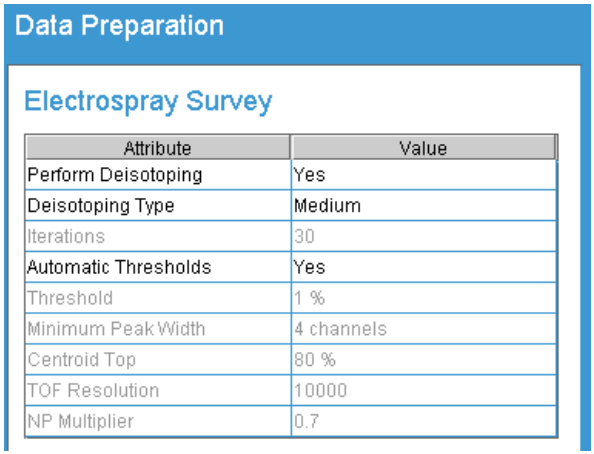
Noise Reduction attributes – MSMS panel:



Attribute	Value
Background Subtract Type	Normal
Background Threshold	35.0 %
Background Polynomial	5
Perform Smoothing	Yes
Smoothing Type	Savitzky-Golay
Smoothing Iterations	2
Smoothing Window	3 channels
Peptide Filter	

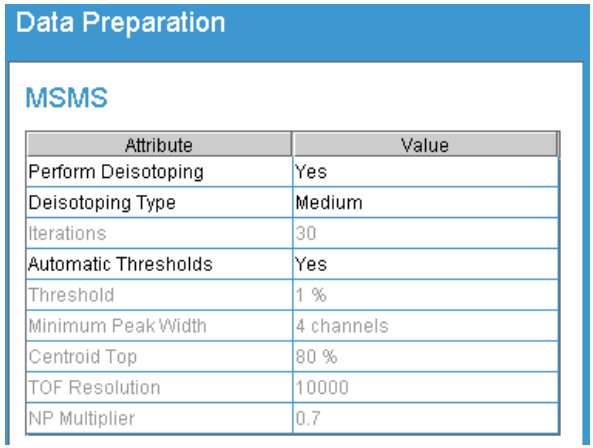
8. Set the Deisotoping and Centroiding attributes in the Electrospray Survey and MSMS panels, as shown below.

Deisotoping and Centroiding attributes – Electrospray Survey panel:



Attribute	Value
Perform Deisotoping	Yes
Deisotoping Type	Medium
Iterations	30
Automatic Thresholds	Yes
Threshold	1 %
Minimum Peak Width	4 channels
Centroid Top	80 %
TOF Resolution	10000
NP Multiplier	0.7

Deisotoping and Centroiding attributes – MSMS panel:



Attribute	Value
Perform Deisotoping	Yes
Deisotoping Type	Medium
Iterations	30
Automatic Thresholds	Yes
Threshold	1 %
Minimum Peak Width	4 channels
Centroid Top	80 %
TOF Resolution	10000
NP Multiplier	0.7

9. Click File > Save As. Save with the file name "Data prep <*current date*>".

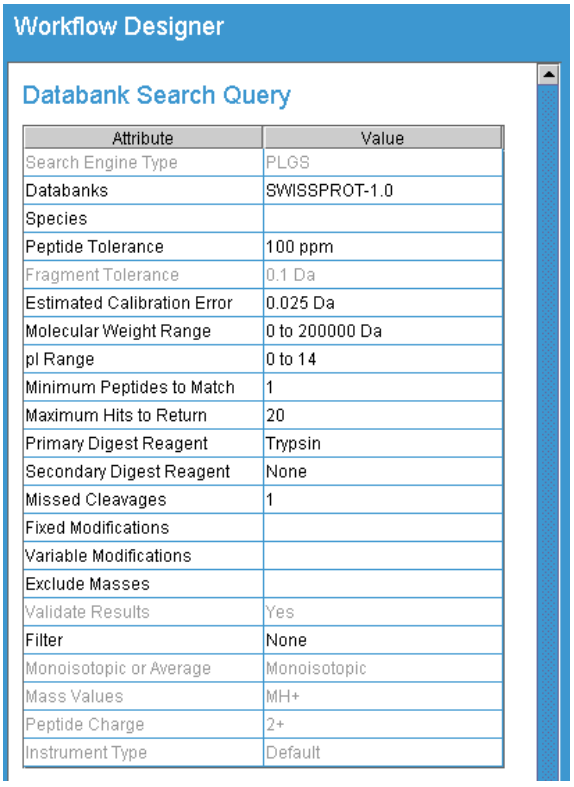
Creating a workflow

To create a workflow:

1. Click Workflow Designer in the tool tray (see Chapter 7 – Defining templates for searching with Workflow Designer).
2. Click File > New.

3. Select Fragment Ion, and then click .
4. Type a title for the workflow (Workflow <*date*>, for example).
5. Right-click the workflow node in the workflow frame, and then click Add > Databank Search.
6. Set the parameters, as shown below.

Databank Search Query parameters:



Attribute	Value
Search Engine Type	PLGS
Databanks	SWISSPROT-1.0
Species	
Peptide Tolerance	100 ppm
Fragment Tolerance	0.1 Da
Estimated Calibration Error	0.025 Da
Molecular Weight Range	0 to 200000 Da
pI Range	0 to 14
Minimum Peptides to Match	1
Maximum Hits to Return	20
Primary Digest Reagent	Trypsin
Secondary Digest Reagent	None
Missed Cleavages	1
Fixed Modifications	
Variable Modifications	
Exclude Masses	
Validate Results	Yes
Filter	None
Monoisotopic or Average	Monoisotopic
Mass Values	MH+
Peptide Charge	2+
Instrument Type	Default

7. Click File > Save As. Save the workflow as "Workflow <*date*>".

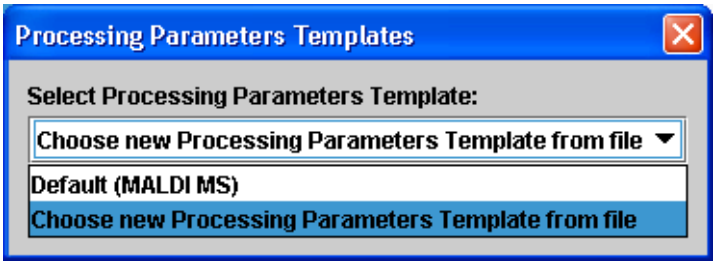
Attaching the data processing parameters

To attach the data processing parameters:

1. In Container Manager, expand the navigator tree so that the Default processing parameters node, directly below the target plate name, is displayed (see Adding processing parameters templates on page 5-21).

2. Click, and then right-click, the Default node
3. Click Change Processing Parameters.

Processing Parameters Templates dialog box:


Processing Parameters Templates
Select Processing Parameters Template:
Choose new Processing Parameters Template from file
Default (MALDI MS)
Choose new Processing Parameters Template from file

4. Click Choose new processing parameters template from file, and then click OK.
5. Click the processing parameter file, Data prep <date>.xml, that you created earlier (see Setting processing parameters on page A-14), and then click Open.

Attaching the workflow file

To attach the workflow file:

1. In Container Manager, highlight all the wells on the plate for which you have set samples (see Setting the target plate on page A-5), by dragging a rectangle over them. Right-click.
2. Select Set Attached Templates > Workflow Template to Mass Spectrum.
3. Click OK, to Choose a new Workflow Template from file.
4. Click the Q-Tof workflow file, Workflow <date>.xml, that you created earlier (see Creating a workflow on page A-17), and then click Open.

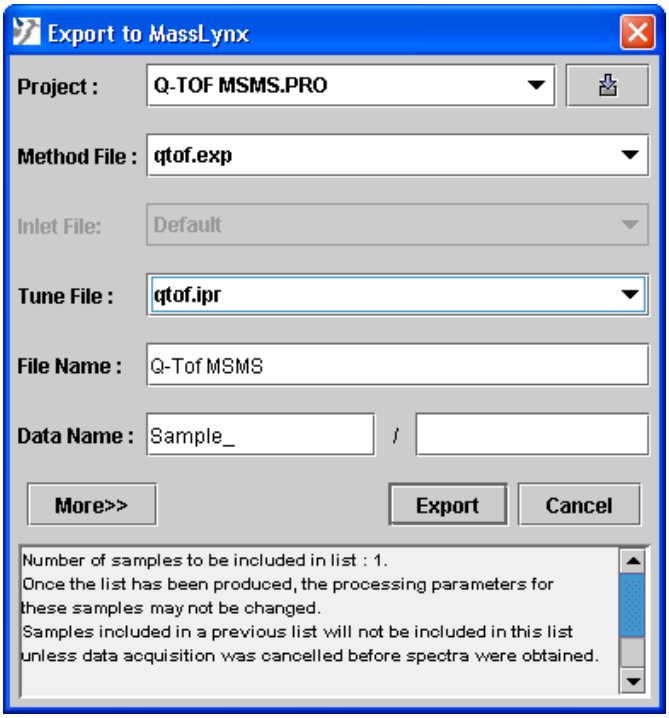
Exporting the sample list to MassLynx

For further details see Exporting a sample list to MassLynx on page 5-29.

To export the sample list:

1. In Container Manager, right-click on the target plate node, and then click Export Sample List to MassLynx.

Export to MassLynx dialog box:


Export to MassLynx
Project :
Q-TOF MSMS.PRO
Method File :
qtof.exp
Inlet File:
Default
Tune File :
qtof.ipr
File Name :
Q-Tof MSMS
Data Name :
Sample_
/
More>>
Export
Cancel
Number of samples to be included in list : 1.
Once the list has been produced, the processing parameters for these samples may not be changed.
Samples included in a previous list will not be included in this list unless data acquisition was cancelled before spectra were obtained.

2. Specify the MassLynx project from which the data is to be acquired.
3. If more than one MS Method is stored in the MassLynx project, use the drop-down list to specify the correct file.

   Tips:

   - The File name can be the same as the target plate name.
   - The MS Data name can be changed to any text, such as digest_0, adh_0.
4. Click Export.
5. In MassLynx click File > Import Worksheet.
6. The file created by PLGS is stored in the MassLynx project. Browse to the file, and then click Open.

Result: The MassLynx sample list is updated with the information from PLGS. Data can now be acquired in the normal way.

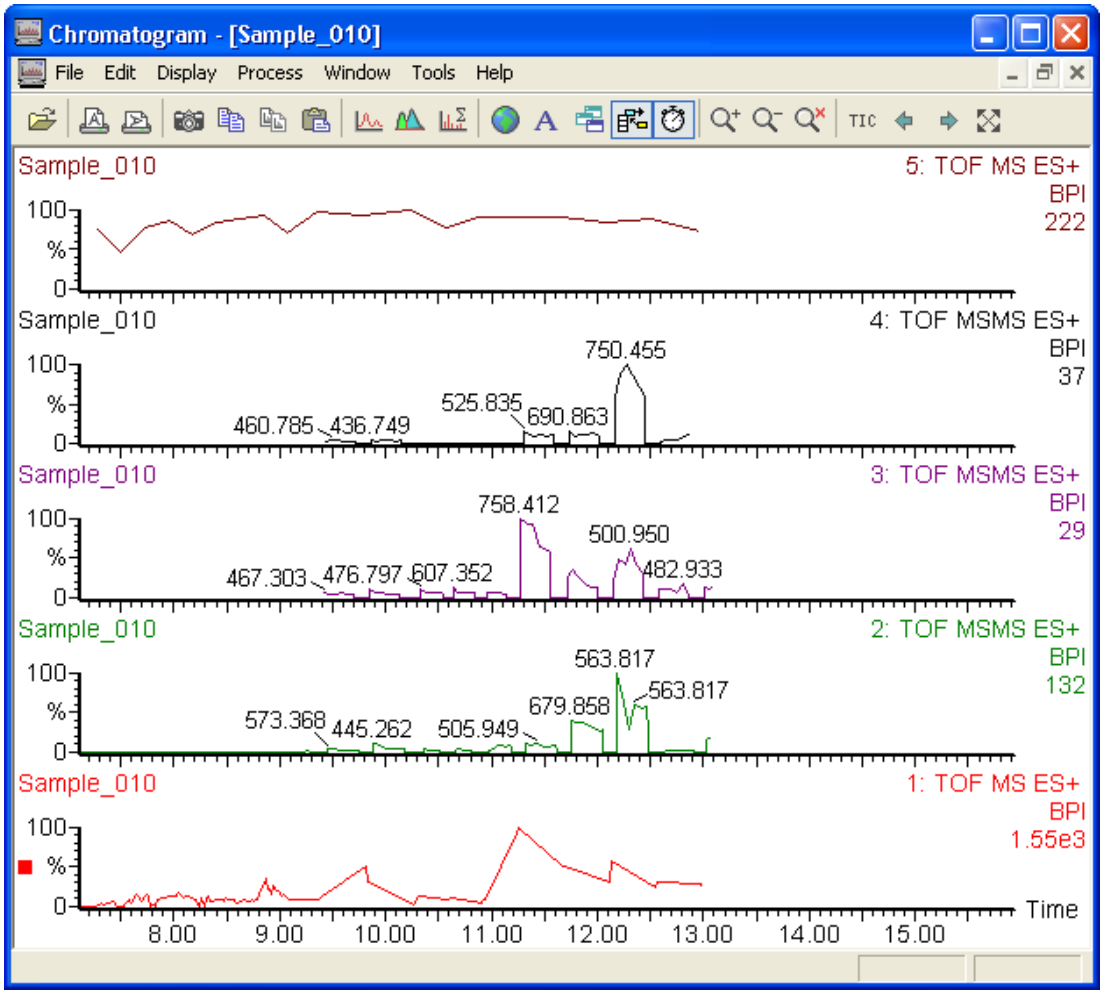
Acquiring data

As the instrument begins to acquire data, chromatograms are recorded. MS data, MSMS data and lockmass correction data is also obtained. When the instrument switches into MSMS mode, the ions selected for MSMS are displayed in the Data Directed Analysis Status.

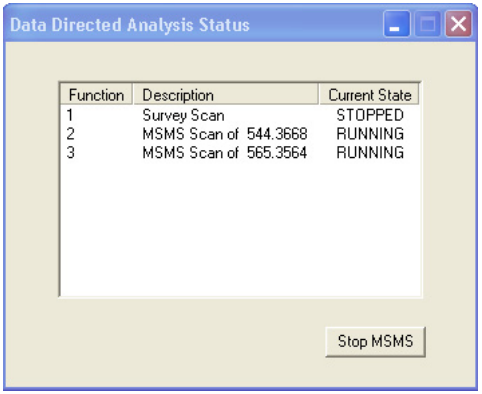
To acquire data:

1. In the main MassLynx window, click to open the Start Sample List Run dialog box.
2. Select Acquire Sample Data and Auto Process Samples.
3. Click OK.

Data Directed Analysis – chromatogram displays:



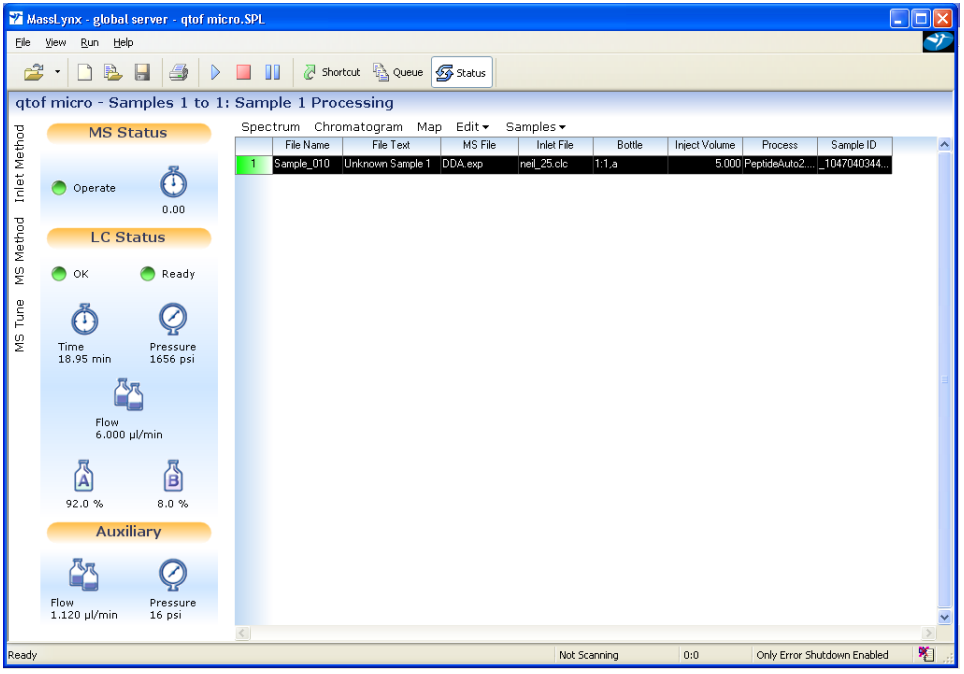
Data Directed Analysis Status display:



At the end of data acquisition Peptide Auto begins processing data information. This is displayed in the PeptideAuto Server window (see Figure titled “PeptideAuto Server display:” on page A-12).

The MassLynx sample list page shows the status of the instrument.

Instrument status in MassLynx:

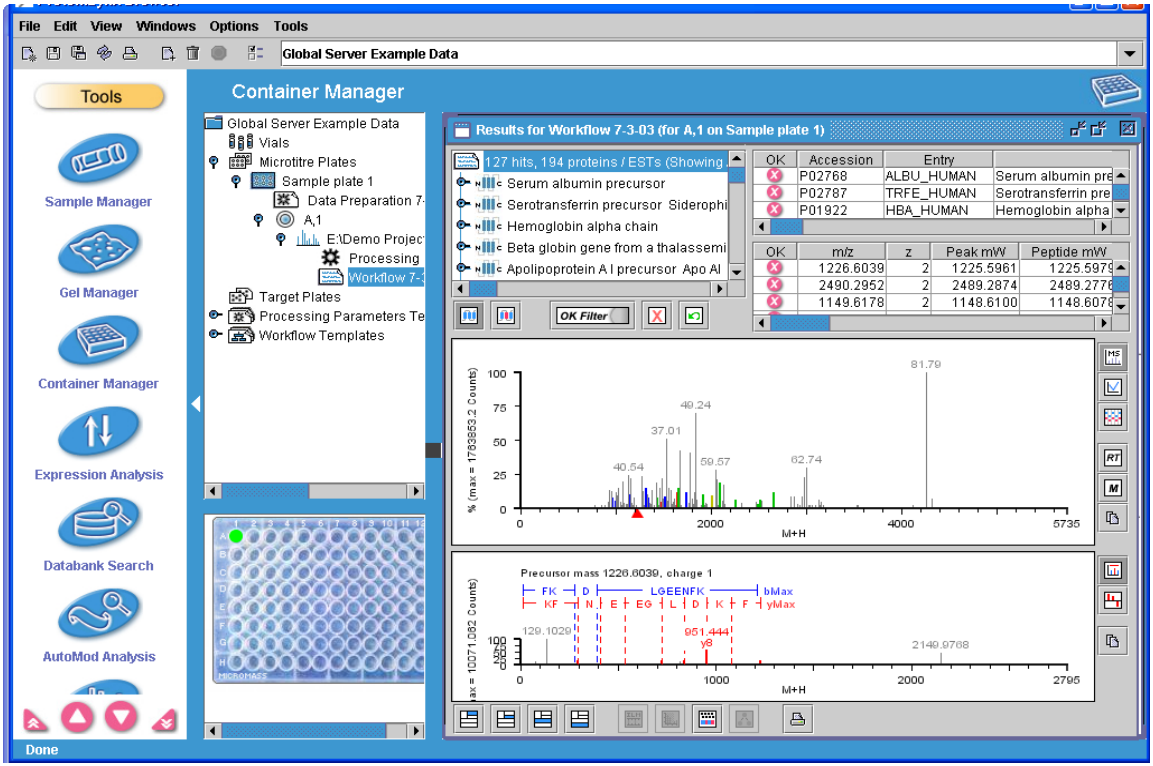


PLGS data processing consists of two major steps:

- Processing MS data, lock mass correcting, and generating lists of precursor mass and charge state.
- Processing the MSMS data, again lock mass correcting and deisotoping data.

When the sample data has been processed and searched against the database, the display in PLGS can be updated. To update the display for the current project in PLGS, click File > Update.

PLGS with acquired data:



Adding a new databank

For further information see Getting started with the Databank Admin tool on page 13-2.

To add a new databank:

1. Click Databank Admin Tool.
2. Click Databanks, and then right-click.
3. Click New Databank.
4. Type a name to use for the databank.
5. Set the following fields:
   - Type to Protein.
   - FASTA Format to, 'STANDARD_SPACED' for Swiss-Prot, or 'NCBI_EXPASY_STANDARD' for the non-redundant database (nrDB).

     See also: Details of the correct format for each database are given in Appendix E, Databanks – Formats.
   - Location, click File and browse to the location of the uncompressed FASTA file on disk - local or mapped.
   - Make Blastable to FALSE - this option creates a BLAST (Basic Local Alignment Search Tool) compatible copy of the database on disk and is required only when sequence data is available.
   - Load into Memory to TRUE if sufficient RAM is available.

     Tip: PLGS can read databases from disk.
   - Management Options to FALSE.
6. Click File > Save Databank Options.

The new database is now available for searching from the client PC.

See also: The download location for nrDB is ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.Z

B Scoring Schemes

This section introduces you to scoring schemes used by ProteinLynx Global SERVER.

Contents:

Topic	Page
Scoring summary	B-2
MALDI scoring (PMF, PMF + fragment ion searches)	B-4
MSMS scoring (fragment ion searches)	B-5
How do I know if a hit is real?	B-6
Automatic data curation	B-7

Scoring summary

The factors contributing to the database search scores are:

- The number of entries in the database – the correct protein(s) are assumed to be in the database and the available probability is initially apportioned equally to each entry in the database.
- When comparing calculated peptide or fragment masses with the data, it is important to know how well the masses in the data are determined. If this estimate is good, the information that can be extracted from the data is maximized. A good estimate increases the scores of correct identifications. An estimate of the precision with which a strong peak can be measured after the instrument is calibrated is the Estimated Calibration Error. There is a further contribution to the overall error estimate that is automatically provided by the de-isotoping software. This further contribution can be significant for weak peaks. The instrument calibration software provides a 'Mean Residual'. To convert this to an estimate of the calibration error, it is recommended that this value is increased by a factor of 1.3.
- Peak area – The importance of a peak is estimated as a function of the signal/noise ratio:

  $$Importance = R\,(Area\ /\ Standard\ Deviation\ of\ Area)^2$$

  Where R is a constant that represents the reliability of detected counts. This gives a measure of the probability of the peak being 'real' as opposed to representing chemical or instrumental noise.
- The number of matched and unmatched peptides – a score is calculated for every peptide in the database. The initial (prior) probability that any given protein in the database is responsible for the submitted data is up or down-rated according to these scores. The scores are reported as natural logs for presentation purposes.
- Fragmentation data – the fragmentation characteristics of peptides at low energy are encoded into a Markov model which incorporates a, b, y, and z immonium ions, fragment ions from modifications, and internal ions from proline. For each peptide sequence, the probability of fragment spectrum GIVEN peptide sequence is calculated. The natural log of this likelihood is the peptide score.

- Search parameters – digest reagent with number of missed cleavages, fixed and variable modifications. Each peptide in each protein in the database is given a prior probability, the weight of which is determined by its end amino acid, the number of missed cleavages it contains, and number of variable modifications it has undergone.

MALDI scoring (PMF, PMF + fragment ion searches)

The scoring scheme implemented in PLGS 2.2.5, for MALDI data, gives a quantitative answer to the question:

“Which single protein best accounts for the data given some initial assumptions?”

The data consists of a set of (mass, intensity) pairs (and their associated uncertainties) representing the mono-isotopic mass and intensity of every peak in the processed data above 900 Da.

All matches (inside the user-set tolerance) are recorded, and ranked according to the scoring scheme.

The reported score, indicating how much of the total probability a protein has, is given by:

$$\textit{Protein Score} = 1n\left(\frac{\textit{Probability of Protein GIVEN Data AND Initial Assumptions)}}{\textit{(Probability of Protein GIVEN Initial Assumption)}}\right)$$

If there are N proteins in the databank, each protein in a databank has a prior probability of '1/N'. Therefore, the maximum possible score is ‘ln(N)’ and the minimum possible top score is zero when the data provides no information, relative to the databank.

The posterior probability of Protein GIVEN Data AND Initial Assumptions is also presented as a percentage.

MSMS scoring (fragment ion searches)

The scoring scheme implemented for MSMS searches addresses the question:

"What is the probability that a protein is in the mixture of proteins that constitutes the sample?"

For this reason there can be more than one hit reported as having maximum (or near maximum) probability of being correct.

The data consists of a set of (mass, intensity) pairs (and their associated uncertainties) representing the mono-isotopic mass and intensity of every peak in the processed data.

- For each precursor ion, a set of peptide sequences is constructed by synthetic digestion of the protein sequences in the database, which match within the user-defined peptide tolerance of the precursor mass.
- For each peptide sequence, the probability of fragment spectrum GIVEN peptide sequence is calculated. The natural log of this is the peptide score. From these probabilities a list is compiled of the most likely combinations of proteins that could have given rise to the data. For example, if we have three proteins, there are 8 possible combinations. The probability of the whole dataset is then calculated given each of these combinations and the probability for a particular protein is accumulated whenever it appears in a combination. We assume that the prior probability of each combination is related to the number of proteins in it and use Bayes' theorem to calculate the probability of protein present in mixture GIVEN dataset. The results are normalized to reflect the number of protein sequences considered in the search.

Therefore:

$$\textit{Probability of A in mixture GIVEN dataset} = \frac{\textit{(SUM over Probabilities of Combinations Containing A)}}{\textit{(SUM over Probabilities of all Combinations)}}$$

Only the highest scoring peptide match is reported for each submitted precursor ion and its associated fragmentation data. Where more than one peptide matches the data equally well (for example if two peptide matches differ only by one or more isobaric residues), all are reported.

How do I know if a hit is real?

To determine if a hit is real, always look to the top scoring protein. Look at the spread of scores: if the scores are grouped together, they will have the same share of the available probability.

In practice, given the variable quality of data, the difference between the top score and the next highest score is usually a good indicator of the correctness of the highest scoring protein. A difference of five (factor of ~ 150) is normally sufficient to indicate that the top scoring protein is correct. Alternatively, a proportion of the available probability can be assumed to be significant, for example, 95%.

For a database with 100,000 entries, the maximum score would be 11.51 and the corresponding '95% significance threshold' would be 11.46 (ln(100,000) + ln(0.95)).

A difficulty can arise with the above criteria when a collection of largely homologous proteins get the top scores. The available probability is then shared between them, for example if the database of 100,000 entries contained two identical sequences that matched the data more closely than any other candidate sequences, the highest scores would approach 10.81 (that is, ln(100,000) + ln(0.5)). In this case, the ProteinLynx browser would present the proteins as a 'collapsed hit', but in other cases it might not be so easy to judge the effective equivalence of the top scoring matches.

To uncover minor components in a sample which contained a mixture of proteins, it is generally not sufficient to read down the list of top scoring proteins, as many of the peptide matches could overlap. It is more appropriate to resubmit the data for searching excluding the top hit. This effectively down-weights data that are matched well by the top hit, which allows independent proteins to score highly.

Other points to consider are:

- As the natural log of values less than 1 results in a negative number, very low scores will be reported as negative numbers in the hit list.
- If the protein being analyzed is not represented (nor has any homologues) in the database, the reported scores will be low and of similar magnitude.
- If a species-specific subset of the database is searched, the scores will be expressed relative to the number of proteins in the subset, rather than the entire database.

Automatic data curation

Depending on the type of search and the search engine used – PLGS or MASCOT – ProteinLynx Global SERVER automatically helps you to organize (curate) your data.

See also: The meanings of 'identity threshold' and 'homology threshold' in relation to the MASCOT search engine are discussed on the Matrix Science website, www.matrixscience.com.

PMF

Automatic data curation rules:

Search engine	Auto-curation?	Requirements for 'OK' assignment	Requirements for 'Maybe' assignment
PLGS	No		
MASCOT	Yes (proteins)	95% identity threshold	Not provided

PMF + Fragment Ion

Automatic data curation rules:

Search engine	Auto-curation?	Requirements for 'OK' assignment	Requirements for 'Maybe' assignment
PLGS	No		
MASCOT	Yes (proteins)	95% identity threshold	Homology threshold

Fragment Ion

Automatic data curation rules:

Search engine	Auto-curation?	Requirements for 'OK' assignment	Requirements for 'Maybe' assignment
PLGS	Yes (if "Validate Results" search parameter set)	All assigned OK	Not applicable
MASCOT	Yes (proteins)	95% identity threshold	Homology threshold

Electrospray-MS

Automatic data curation rules:

Search engine	Auto-curation?	Requirements for 'OK' assignment	Requirements for 'Maybe' assignment
PLGS	Yes	95% probability	50% probability
MASCOT	Yes	95% identity threshold	Homology threshold

Electrospray-High/Low

Automatic data curation rules:

Search engine	Auto-curation?	Requirements for 'OK' assignment	Requirements for 'Maybe' assignment
PLGS	Yes	95% probability	50% probability
MASCOT	Yes	95% identity threshold	Homology threshold

C Implementing a plugin for ProteinLynx Global SERVER

This section provides PLGS users with an overview of the plugin system used within the PLGS applications. After reading this section you should understand the plugin architecture that exists within PLGS and also have an appreciation of how you can design and create your own custom plugins, which can then be used within PLGS.

Contents:

Topic	Page
An introduction to the PLGS plugin	C-2
Plugin architecture	C-3
Use case – the PLGS FileSystemPlugIn	C-5
XML communication with the plugin implementation	C-6
Adding a plugin to the PLGS application	C-7
An example Executable plugin	C-11
An example Java plugin	C-13
Basic plugin-Specific Queries	C-16
Query tag definitions in the ProteinLynx DTD	C-21
Plugin process exit codes	C-26
UML Class Diagram for the PLGS plugin Architecture	C-27

An introduction to the PLGS plugin

A plugin can be thought of as a means to 'plug in' to a system or application and allow for the transfer of data in to or out of that system.

Since PLGS 2.0 the PLGS applications have utilized plugins. The default plugin used within PLGS is a simple plugin that allows data to be imported and exported from an underlying file system to the ProteinLynx Browser. This plugin has thus been termed the "FileSystemPlugIn".

Every time you press the save button in the browser, a request is sent to the default plugin to take the associated data and to store it appropriately in the underlying file system. Similarly, when you select to import data into the browser (such as a databank search), another request is sent to the plugin to find the associated data within the underlying file system and to return this data to the browser for display.

The FileSystemPlugIn is the default plugin used within PLGS, but you might wish to design and create custom plugins in order to handle PLGS data in a custom manner. In order to do this we must further explore the architecture of plugins.

Plugin architecture

Plugins can be implemented in any programming language that allows access to the standard data streams. For example, a C language implementation would receive its input through 'stdin' and provide output through 'stdout'. Any error messages would be channeled through 'stderr'. The integer return value of the main function can also be used to signal the exit status from the plugin (see Plugin process exit codes on page C-26).

In order to meet user requirements and integrate with third party databases or LIMS systems, a plugin interface has been designed for PLGS since ProteinLynx 2.0. The plugin interface provides third parties with a means to import or export data into or out of PLGS.

The plugin architecture provides a simple interface to external data sources. A plugin makes a call back to its associated PlugInHandler in a set order after its run() method has been invoked.

- Immediately after the plugin has started, the handleStart() will be called. This method provides the required streams to the handler, input, output and error streams. If input to the plugin is to be provided before being acted upon, it should be written and the stream closed. If a large amount of output is expected it is probably most efficient to perform blocking reads from the output stream until the stream is closed.
- Once the handleStart() method has been called it can be followed by calls to handleOutput() or handleError().
- handleOutput() - will be called when bytes are available form the output stream. If more output is expected this method should return true.
- handleError() - will be called when bytes are available form the error stream. If more output is expected this method should return true.
- Finally, either of handleException() or handleEnd() will be called but not both.
- handleException() - if an exception arises which cannot be dealt with using a status code, handleException() is invoked in place of handleEnd().
- handleEnd() – if the plugin reaches the end of its task, this method is invoked with a status code.

PlugInHandlers can be implemented for specific tasks, although some generic implementations can prove useful – an OutputStreamPlugInHandler, for example.

Currently there are two plugin implementations provided with PLGS, Executable and Java class implementations.

Executable plugins or ExecPlugIns extend the plugin interface to allow executables to be used to import and export data into and out of PLGS.

Java class plugins extend the plugin interface to allow classes which implement an additional interface called the PlugInImp interface to import and export data into and out of PLGS. PlugInImp classes simply process input from the plugin through an input stream, process output from the plugin through an output stream and process error messages from the plugin through an error stream. A UML class diagram of this plugin architecture can be found in UML Class Diagram for the PLGS plugin Architecture on page C-27.

The client of a plugin is the item or entity that calls and runs that plugin. The dialogue required between a client and a plugin is particularly simple: all input is provided by the client and then the input stream is closed, the client of the plugin then waits for output or the termination of the plugin process. All plugins have an associated PlugInHandler that will handle plugin events such as the start of the plugin process, handling output from the plugin; handling errors form the plugin and handling the end of the plugin process.

Use case – the PLGS FileSystemPlugIn

The FileSystemPlugIn is the default import and export plugin used by PLGS. It is used in order to save (import) data into a PLGS project and also to retrieve data from (export) a PLGS project held on an underlying file system structure. The file system structure consists of the following:

- Root Directory (Project Store)
- Project Folder
- Sample Tracking Folders
- Workflow Results Folder for Parent Sample Tracking
- Gels Folder
- Expression Analyses Folder
- Expression Analysis Folders
- Expression Analysis Results Folder for Parent Expression Analysis

The FileSystemPlugIn is a Java class plugin, it extends the PlugInImp interface. This means that the FileSystemPlugIn has 2 distinct methods –setProperties() and process().

The setProperties() method is used to set specific properties for the FileSystemPlugIn and is called immediately after the FileSystemPlugIn is instantiated.

The process() method is used to process the input, output and error messages from the FileSystemPlugIn. The input is read from the input stream, while the output is written to the output stream, and error output is directed to the error stream.

After the FileSystemPlugIn has been instantiated and its properties have been set, it is assigned a PlugInHandler. This handler defines how the individual plugin events should be handled.

XML communication with the plugin implementation

In order to allow easy integration with third party systems, communication between the data storage system and the ProteinLynx system, an XML-based query language is defined in the ProteinLynx Document Type Definition (DTD). PLGS can communicate with a plugin by way of a series of predefined XML queries. There are a series of query types:

- Select
- Insert
- Delete
- Update

Within the DTD, a set of elements related to querying XML and other documents is specified. These elements constitute a primitive query language. Essentially, the Project, Workflow and Mass Spectrum XML documents, described in the DTD, along with gel images, sample lists, and Expression Analysis experiments are the blocks of data by which the ProteinLynx system communicates. For examples of the types of plugin specific queries, see Basic plugin-Specific Queries on page C-16.

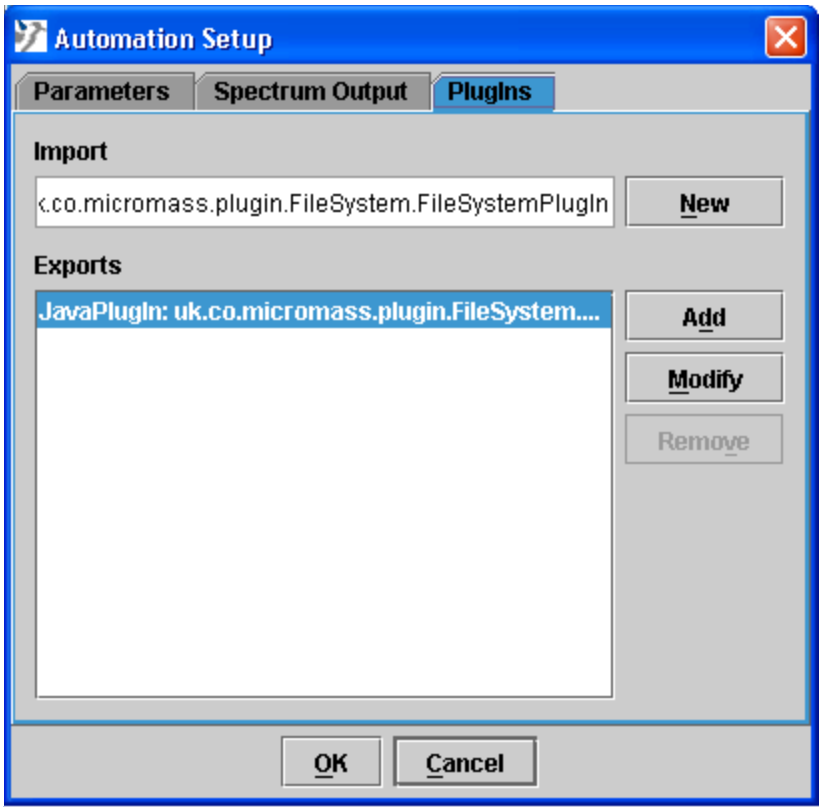
Adding a plugin to the PLGS application

Once a new plugin has been created it needs to be added to the list of plugins in PLGS.

To add a plugin:

1. Start the browser.
2. Click Options > Automation Setup.
3. Click the PlugIns tab.

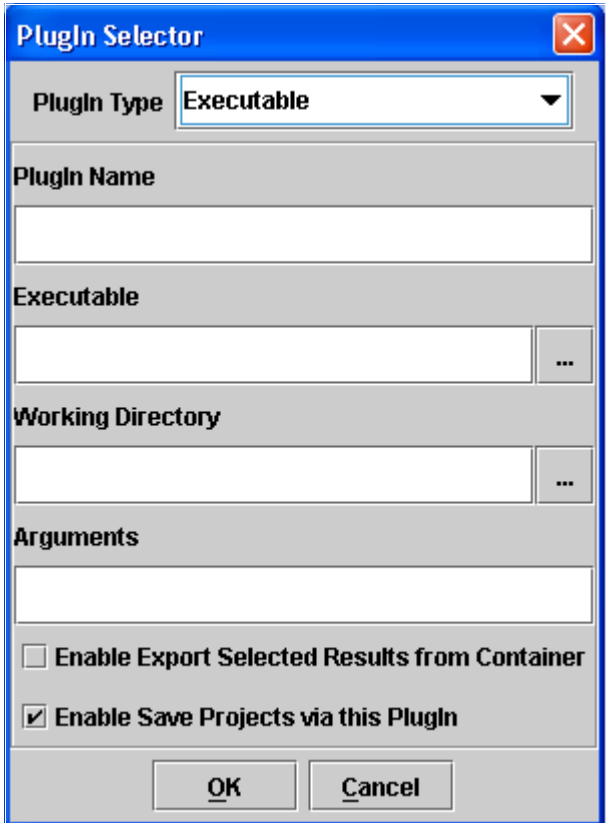
Automation Setup dialog box - PlugIns tab:



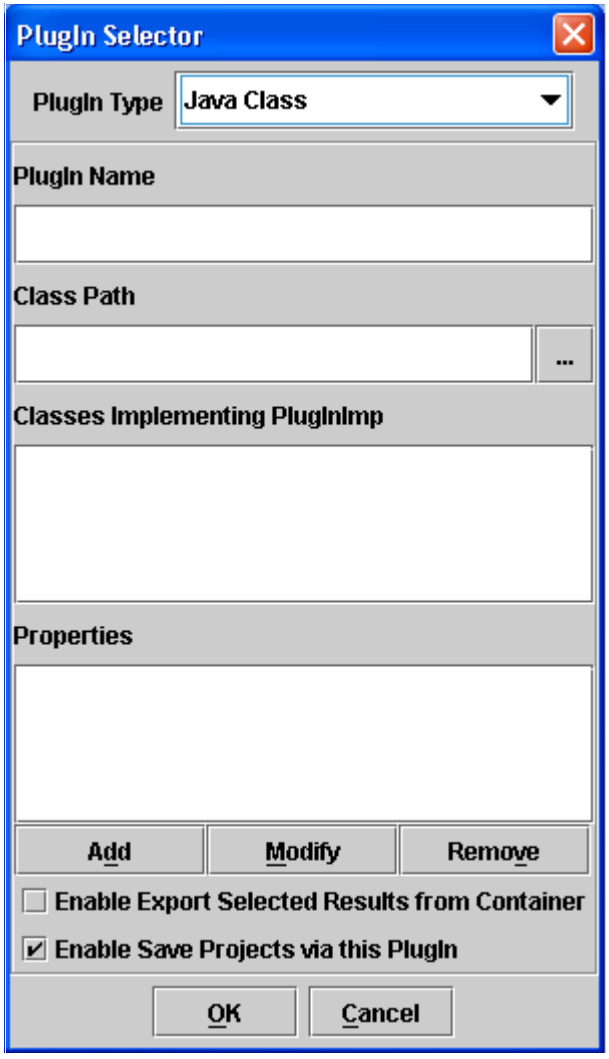
4. Click Add.

   The PlugIn Selector dialog box opens, in which you can set up either an Executable or Java Class type of plugin.

PlugIn Selector dialog box - Executable plugin type:

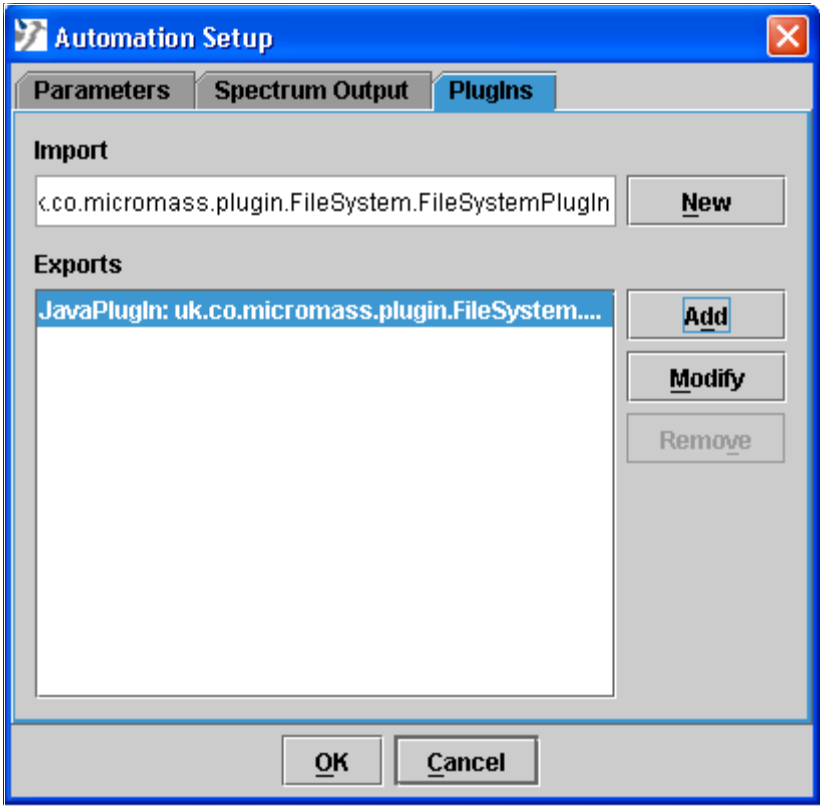


PlugIn Selector dialog box - Java Class plugin type:



5. Select either an Executable or Java Class type of plugin and set the parameters.
6. Once added successfully the new plugin is displayed in the Exports list.

PlugIns page - Plugin displayed in Exports list:



When an item is saved it will be passed to the new plugin as well as to the default FileSystemPlugIn.

An example Executable plugin

The following is the source required to create an example Executable plugin called HelloPlugIn.exe.

Build this code in Visual Studio to create the executable and then add it to the exports in PLGS. The HelloPlugIn.exe takes the input to the plugin and then prints it out to a file called helloplugin1.txt, which can be found in the working directory you set when adding the plugin to the list of export plugins. Try this and see how it works.

```
// HelloPlugin1
// Reads input from stdin and writes it to file
#include <fstream>
#include <iostream>
#include <string>
using namespace std;

int main( int argc, char* argv[] )
{
      ofstream out;
      // file to write input to
      out.open( "helloplugin1.txt", ios::app );
      // ensure file has opened
      if ( !out )
      {
           cerr << "HelloPlugin1 - ERROR OPENING helloplugin1.txt!" << endl;
            return 3;
      }

    while ( cin )
      {
```

```
            cin.getline(c);
            out << c;
        }
        out<<endl;
        // close file
        out.close();
        // return SUCCESS exit code
        return 0;
}
```

An example Java plugin

All Java class plugins must implement this interface to become compatible with PLGS.

The PlugInImp interface has 2 methods:

```
/**
 * Processes the input read from input stream, writing output to output stream. Error output is directed to error stream.
 * @param inputStream the input stream
 * @param outputStream the output stream
 * @param errorStream the error stream
 * @return 0 for success
 * @exception java.lang.exception when the processing cannot continue due to an error
 */
 int process(java.io.InputStream inputStream, java.io.OutputStream outputStream, java.io.OutputStream errorStream) throws java.lang.Exception;


/**
 * Sets properties for this PlugInImp. Called immediately after the PlugInImp is instantiated.
 * @param properties the properties for this PlugInImp
 * @exception implementations should throw an IllegalArgumentException if necessary properties are absent or invalid
*/
 void setProperties(java.util.Properties properties);
```

The following is the source code for an example Java class plugin called MirrorPlugIn.java. This plugin will print out the input it receives to the System.out. Notice how this class implements the PlugInImp interface. Add this plugin in PLGS to see how it works. In order for the MirrorPlugIn to become available you must compile it and place the MirrorPlugIn.class into a jar file with the PlugInImp.class which can be found in the proteinprobe.jar file in the PLGS installation folder called “jars”.

```
/*
* Created on 26-Sep-2003
*/
package MirrorPlugIn;
import java.io.InputStream;
import java.io.InputStreamReader;
import java.io.OutputStream;
import java.util.Properties;
import uk.co.micromass.plugin.PlugInImp;
/**
 * @author NEESONK
 *
 * To change the template for this generated type comment go to
 * Window&gt;Preferences&gt;Java&gt;Code Generation&gt;Code and Comments
 */
public class MirrorPlugIn implements PlugInImp
{
    /**
    * This is the main process method of all Java plugins
    */
    public int process(InputStream inputStream, OutputStream
    outputStream, OutputStream errorStream) throws Exception
    {
```

```
        System.out.println( "The MirrorPlugIn has been called" );
        InputStreamReader reader = new InputStreamReader( inputStream );
        char [] buf = new char[1024];
        int nRead = 0;
        StringBuffer buffer = new StringBuffer();

        System.out.println( "Here comes the input to the MirrorPlugIn" );
        do

        {

            nRead = reader.read( buf, 0, buf.length );
            if (nRead > 0)

            {

                //os.write(Buf, 0, nRead);
                buffer.append(buf, 0, nRead );

            }
     } while( nRead != -1 );

        System.out.println(buffer.toString());
        System.out.println( "The MirrorPlugIn has finished" );
        return 0;

    }
    /**

     * This method is used to set any properties the PlugIn may have

     */

    public void setProperties(Properties properties)

    {

        // To do: Auto-generated method stub.

    }
}
```

Basic plugin-Specific Queries

There are four basic plugin-specific queries:

- Selection of elements
- Update of elements
- Deletion of elements
- Insertion of documents

Selection of elements

```
<?xml version="1.0" ?>
<QUERY>
    <SELECT ELEMENT_TYPE=" PROJECT " RETURN="document">
        <REFERENCE NAME="PROJECT">
                <REF_ATTRIBUTE NAME="PROJECT_ID"
                VALUE="Project3" />
        </REFERENCE>
    </SELECT>
</QUERY>
```

Selecting a Project document for a given Project ID

Above is an example query to the FileSystemPlugIn. The query is asking the plugin to select the Project with the Project ID of Project 3. This example clearly illustrates how simple queries can be built.

All queries have an outer <QUERY> tag and within this tag will be a series of descriptive elements to define the query. In this instance the query action is a SELECT and thus a select element has been inserted which describes the type of document to select and what format the returned document should be in. In this case the entire document is returned as opposed to a URL of the documents location.

In a returned QUERY element, a list of references can express the results of the query. In the case of large documents (usually MASS_SPECTRUM documents containing fragmentation data), it can be more efficient to return a URL to the document than to stream the document directly through the plugin. The return attribute of the SELECT element allows the client to specify that the plugin return a URL or a reference, rather than a document.

All plugin queries also contain an inner reference element, which provide a reference for the query document. Reference tags have a single NAME attribute and one or more inner <REF_ATTRIBUTE> elements, which help describe particular attributes of the referenced document. In this case the referenced document is a project that has a PROJECT_ID attribute set to "Project 3".

The selection of elements of the specified type is predicated upon them having attributes or child elements with attributes matching all those specified by the given reference tree.

Update of elements

```
<?xml version="1.0" ?>
<QUERY>
    <UPDATE ELEMENT_TYPE=”PROJECT”>
        <REFERENCE NAME=”PROJECT”>
            <REF_ATTRIBUTE NAME=”PROJECT_ID” VALUE=” Project3”/>
        </REFERENCE>
        <TAG>
                <PROJECT …>
                    …
                </PROJECT>
        </TAG>
    </UPDATE >
</QUERY>
```

Updating a Project document for a given Project_ID

Update queries, like select queries, are done at the element level. The insertion or deletion of an element within a document can be thought of as an update to the parent element. Therefore, an update comprises the location of the element to be changed (or the parent element of elements to be deleted or inserted) and the specification of its replacement, if the element has a required attribute of type ID.

As shown in the example above, the descriptive element UPDATE is very similar to the SELECT element in the previous example; note that the REFERENCE element is exactly the same. An update query contains an additional <TAG> element – this element contains the updated version of the item to be updated. This element might, for example, contain an entire Project document: the referenced project would then be located and updated with the updated version.

Deletion of elements

```
<?xml version="1.0" ?>
    <QUERY>
        <DELETE ELEMENT_TYPE=" MASS_SPECTRUM " >
            <REFERENCE NAME="MASS_SPECTRUM">
                <REF_ATTRIBUTE NAME="SAMPLE_TRACKING_ID"
                VALUE="B001" />
            </REFERENCE>
        </DELETE>
    </QUERY>
```

Deleting a Mass Spectrum document for a given Sample Tracking ID

Elements for deletion are selected in the same way as in a select query. The only difference is that the query action is a DELETE rather than a SELECT. Note that there is no return type as no document can be returned after it has been deleted. The above example has selected the Mass Spectrum document for Sample Tracking ID B001 to be deleted.

Insertion of documents

```
<?xml version="1.0" ?>
<QUERY>
    <INSERT>
        <TAG>
            <WORKFLOW …>
                    …
            </WORKFLOW>
        </TAG>
    </INSERT>
    <UPDATE ELEMENT_TYPE=”PROJECT”>
        <REFERENCE NAME=”PROJECT”>
            <REF_ATTRIBUTE NAME=”PROJECT_ID” VALUE=” Project3”/>
        </REFERENCE>
        <TAG>
            <PROJECT …>
                    …
            </PROJECT>
        </TAG>
    </UPDATE>
</QUERY>
```

Inserting a Workflow document and updating the associated Project document

Documents can be inserted either by specifying the entire document or by specifying a URL at which the documents can be found. In the above example a workflow is to be inserted. The entire workflow document is located in the INSERT block and this is then followed by an update query for the Project with the PROJECT_ID - Project 3. Alternatively, a URL can be provided inside a REFRENCE element as illustrated in the following example code.

```
<?xml version="1.0" ?>
<QUERY>
    <INSERT>
        <REFERENCE NAME=”MASS_SPECTRUM”>
            <REF_ATTRIBUTE NAME=”SAMPLE_TRACKING_ID”
            VALUE=”_98375409685408”/>
        </REFERENCE>
        <URL PROTOCOL=”file”
        PATH=”C:/temp/mass_spectrum_27634.xml”/>
    </INSERT >
    <UPDATE ELEMENT_TYPE=”PROJECT”>
        <REFERENCE NAME=”PROJECT”>
            <REF_ATTRIBUTE NAME=”PROJECT_ID” VALUE=” Project3”/>
        </REFERENCE>
        <TAG>
            <PROJECT …>
                    …
            </PROJECT>
        </TAG>
    </UPDATE>
</QUERY>
```

Query tag definitions in the ProteinLynx DTD

Here is the section of the DTD that is specific to plugin activity.

```
<!-- Query

    Description:
    This describes a query to a plugin and possibly the result of the query.

    Documents:
    Query

    Attributes:
    @username - the username for authentication
    @password - the password for authentication

-->

<!ELEMENT QUERY (

    ( ( INSERT | UPDATE | SELECT | DELETE )+ ) | TAG

)>

<!ATTLIST QUERY

    USERNAME                CDATA                #IMPLIED

    PASSWORD                CDATA                #IMPLIED

>

<!-- Insert

    Description:
    Describes a document to be inserted.

    Document:
    Query

    Attributes:

-->

<!ELEMENT INSERT (
    ( REFERENCE , URL ) | TAG

)>

<!-- Update
```

```
    Description:
    Describes an element to be updated.

    Document:
    Query

    Attributes:
    @element-type - the type of element.

-->

<!ELEMENT UPDATE (
    REFERENCE* ,
        TAG

)>

<!ATTLIST UPDATE
    ELEMENT_TYPE                CDATA                   #REQUIRED

>

<!-- Select

    Description:
    Describes a select query.

    Document:
    Query

    Attributes:
    @element-type - the type of element to return
    @return - the type of data to return

-->

<!ELEMENT SELECT (
    REFERENCE*

)>

<!ATTLIST SELECT
    ELEMENT_TYPE                CDATA                   #REQUIRED
    RETURN ( document | reference | url )"document"

>

<!-- Delete

    Description:
    Describes a document to be deleted.
```

```
    Document:
    Query

    Attributes:
    @type - the type of element to be deleted
-->
<!ELEMENT DELETE (
    REFERENCE*
)>
<!ATTLIST DELETE
    ELEMENT_TYPE                CDATA                   #REQUIRED
>
<!-- Tag

    Description:
    A place-holder for any tag.

    Document:
    Query

    Attributes:
-->
<!ELEMENT TAG ANY>
<!-- Reference

    Description:
    This is a reference to an external XML document.
    It is also used to define the element to search for in a query.

    Documents:
    Project, Query

    Attributes:
    @name - the name of the element
-->
<!ELEMENT REFERENCE (
    REF_ATTRIBUTE*,
    REF_TEXT?,
    REFERENCE*
```

```
)>
<!ATTLIST REFERENCE
    NAME                    CDATA                   #REQUIRED
>
<!-- Ref_attribute
    Description:
    This describes an attribute of an element referred to by a reference.
    Documents:
    Query
    Attributes:
    @name - the name of the attribute
    @value - the value of the attribute
-->
<!ELEMENT REF_ATTRIBUTE EMPTY>
<!ATTLIST REF_ATTRIBUTE
    NAME                    CDATA                   #REQUIRED
    VALUE                   CDATA                   #REQUIRED
>
<!-- Ref_text
    Description:
    This describes PCDATA of an element referred to by a reference.
    Documents:
    Query
    Attributes:
-->
<!ELEMENT REF_TEXT ( #PCDATA )>
<!-- Url
    Description:
    Describes a url.
    Documents:
    Query, Project
```

```
    Attributes:
    @protocol - the protocol
    @host - the hostname or ip address
    @port - the port number
    @path - the path

-->

<!ELEMENT URL EMPTY>
<!ATTLIST URL
    PROTOCOL            ( http | https | file )         "file"
    HOST                    CDATA                       #IMPLIED
    PORT                    CDATA                       #IMPLIED
    PATH                    CDATA                       #REQUIRED

>
```

Plugin process exit codes

The plugin process exit codes are:

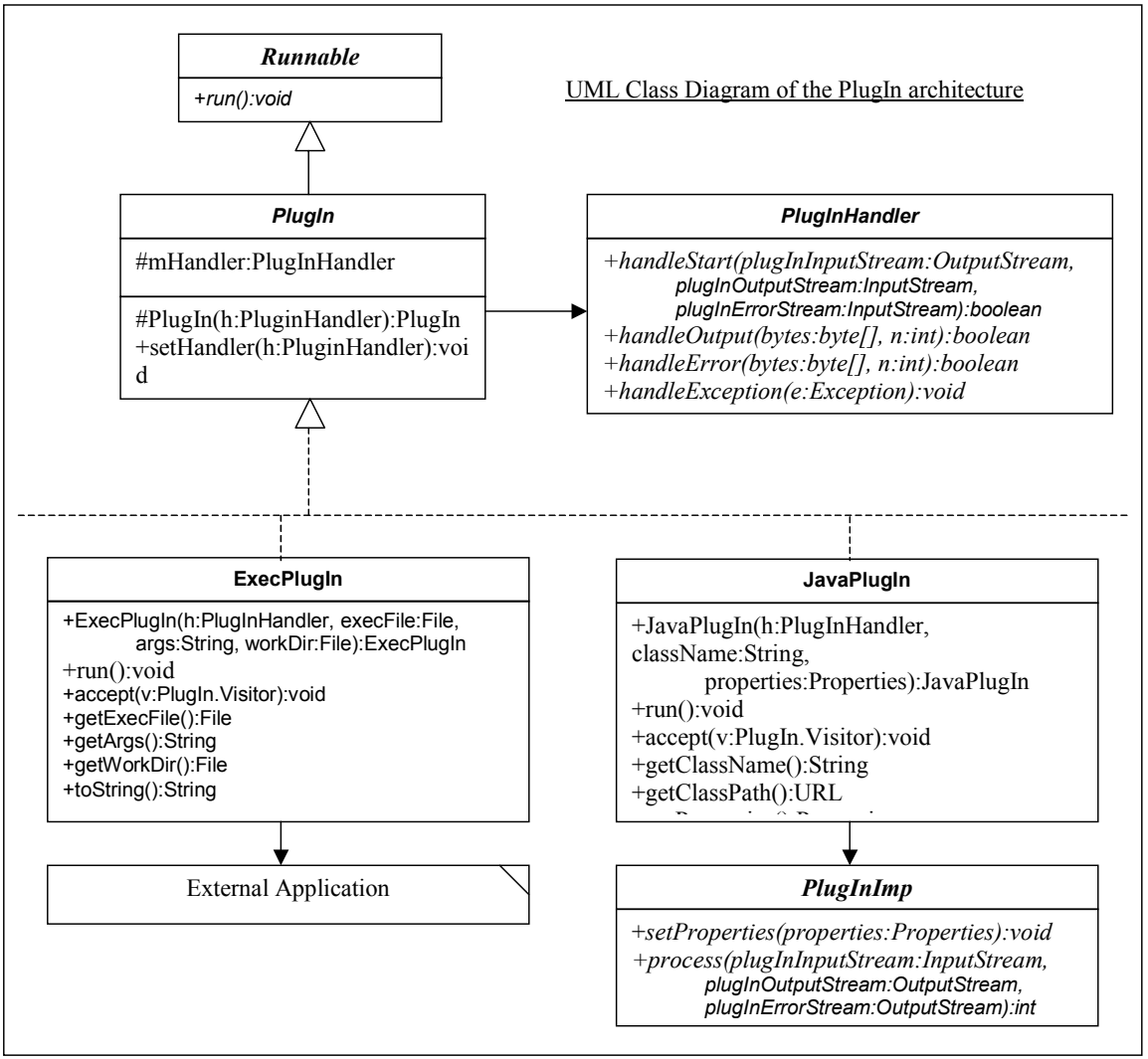
Plugin process exit codes:

Code	Description
0	Successful completion
1	File not found
2	Invalid query
3	Error
4	Busy

UML Class Diagram for the PLGS plugin Architecture

The following diagram illustrates the PLGS plugin architecture.

UML Class diagram for the PLGS plugin architecture:



D UNIX Help for Installing PLGS on AIX Platforms

This section describes using command line input to install PLGS on AIX platforms.

All changes can be made from the command line. In most cases, however, the more user-friendly SYSTEM MANAGEMENT INTERFACE TOOL (SMIT) can be used. SMIT can be invoked from the command line by typing the command SMIT, or by clicking on the Common Desktop Environment. When possible, reference to executing a command through SMIT will be included.

Contents:

Topic	Page
Installing PLGS using the command line	D-2

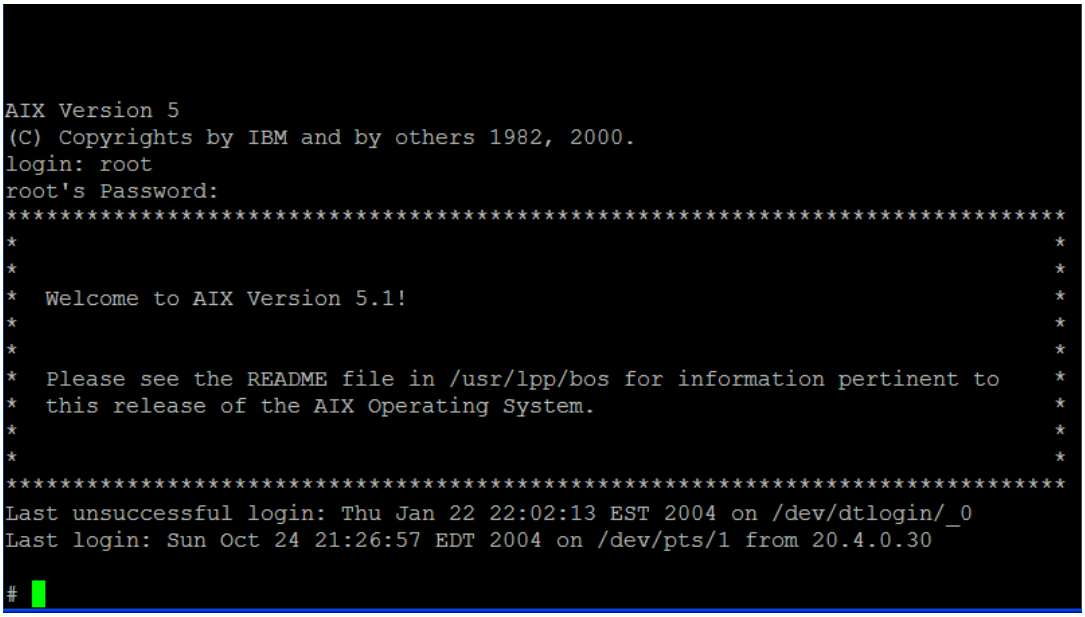
Installing PLGS using the command line

To install PLGS using the command line:

1. Login as root.

   The login window is either a regular command line window or a Common Desktop Environment (a graphical user interface).

   #### Logging in as root:


AIX Version 5
(C) Copyrights by IBM and by others 1982, 2000.
login: root
root's Password:
Welcome to AIX Version 5.1!
Please see the README file in /usr/lpp/bos for information pertinent to
this release of the AIX Operating System.
Last unsuccessful login: Thu Jan 22 22:02:13 EST 2004 on /dev/dtlogin/_0
Last login: Sun Oct 24 21:26:57 EDT 2004 on /dev/pts/1 from 20.4.0.30
#

   In a terminal window, the prompt symbol indicates what shell you are using. The #, $ and & respectively represent the Korn, Bourne and C shells.

2. Check if the TMPDIR variable exists. Setting the TMPDIR creates a pointer to a location where there is sufficient space for working files.

   At the prompt type the command:

   ```
   env | pg
   ```

3. Press Enter.

   The environmental variables are displayed.

Example:

```
TMPDIR=/usr/tmp
myid=dot
LANG=En_US
UNAME=davisd
PAGER=/bin/pg
VISUAL=vi
PATH=/usr/ucb:/usr/lpp/X11/bin:/bin:/usr/bin:/etc:/u/dot:/u/dot/bin:/u/bin1
MAILPATH=/usr/mail/dot?dot has mail !!!
MAILRECORD=/u/dot/.Outmail
EXINIT=set beautify noflash nomesg report=1 showmode showmatch
EDITOR=vi
PSCH=>
HISTFILE=/u/dot/.history
LOGNAME=dot
MAIL=/usr/mail/dot
PS1=dot@davisd:${PWD}>
PS3=#
PS2=>
epath=/usr/bin
USER=dot
SHELL=/bin/ksh
HISTSIZE=500
HOME=/u/dot
FCEDIT=vi
TERM=lft
MAILMSG=**YOU HAVE NEW MAIL. USE THE mail COMMAND TO SEE
YOUR PWD=/u/dot
ENV=/u/dot/.env
```

Adding TMPDIR

To add TMPDIR:

1. Type the commands:

   `TMPDIR=/` (Where ever you have large space allocation on system.)

   `export TMPDIR`

2. Type:

   `env | pg`

   This verifies that the TMPDIR path has been set correctly.

Mounting a CD-ROM

To mount a CD-ROM:

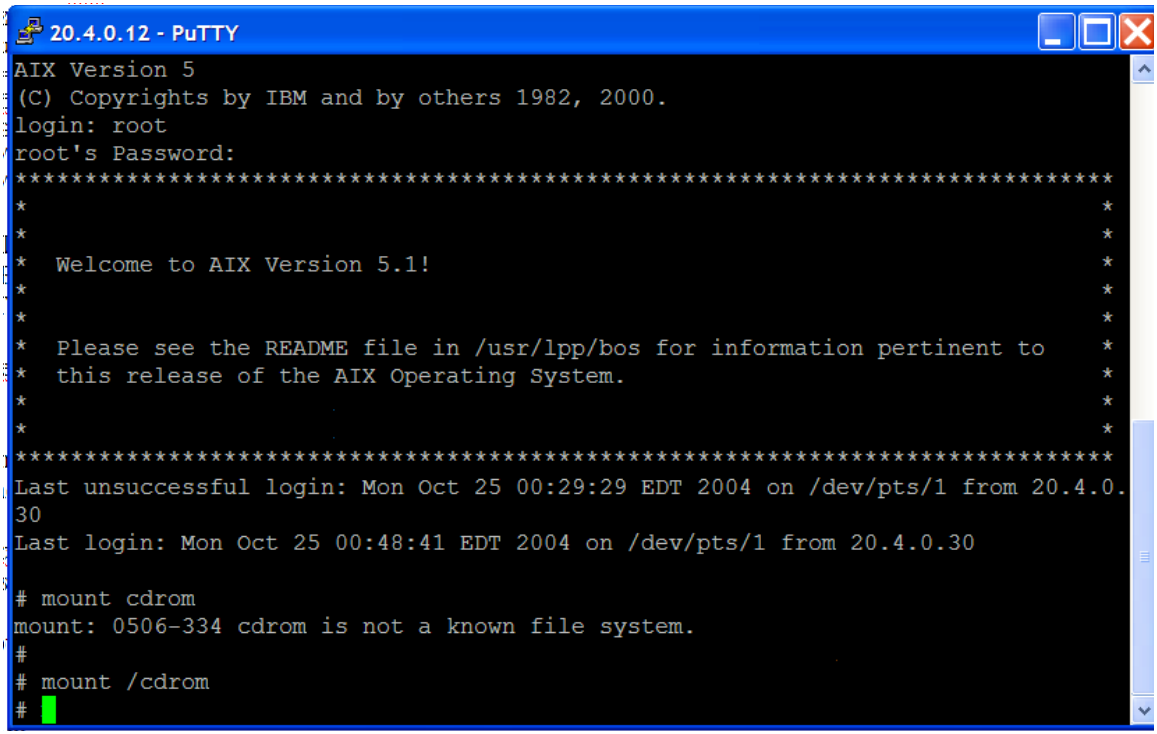
1. Insert the CD, and then at the command prompt type:

   `mount /cdrom`

2. Press Enter.

   This mounts the CD-ROM on the file system cdrom. The CD-ROM drive should spin up. If you type the command incorrectly or omit the / an error will occur.

Mounting a CD-ROM:

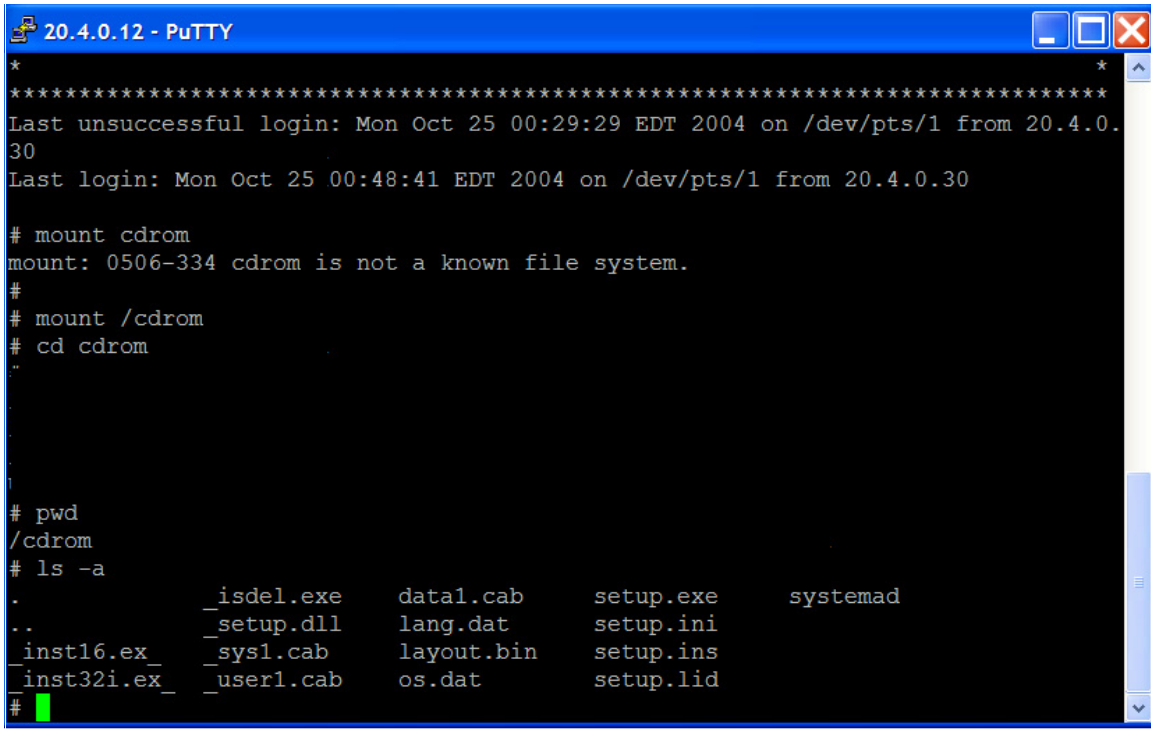

20.4.0.12 - PuTTY
AIX Version 5
(C) Copyrights by IBM and by others 1982, 2000.
login: root
root's Password:
Welcome to AIX Version 5.1!
Please see the README file in /usr/lpp/bos for information pertinent to this release of the AIX Operating System.
Last unsuccessful login: Mon Oct 25 00:29:29 EDT 2004 on /dev/pts/1 from 20.4.0.30
Last login: Mon Oct 25 00:48:41 EDT 2004 on /dev/pts/1 from 20.4.0.30
mount cdrom
mount: 0506-334 cdrom is not a known file system.
#
mount /cdrom
#

3. To verify you have mounted the CD, type the commands:

   ```
   cd /cdrom
   pwd
   ls -a
   ```

   The contents of the CD should be listed.

Listing the contents of a CD-ROM:



Using SMIT

If the CD-ROM does not mount, go to SMIT to check what the CD-ROM drive is referenced as.

To check the CD-ROM drive reference:

1. Open SMIT.
2. Select System Storage Management (Physical & Logical Storage).
3. Select File Systems.
4. Select List All File Systems.
5. In the list locate the device /dev/cd0. The mount point is the reference to be used.
6. Click Done.

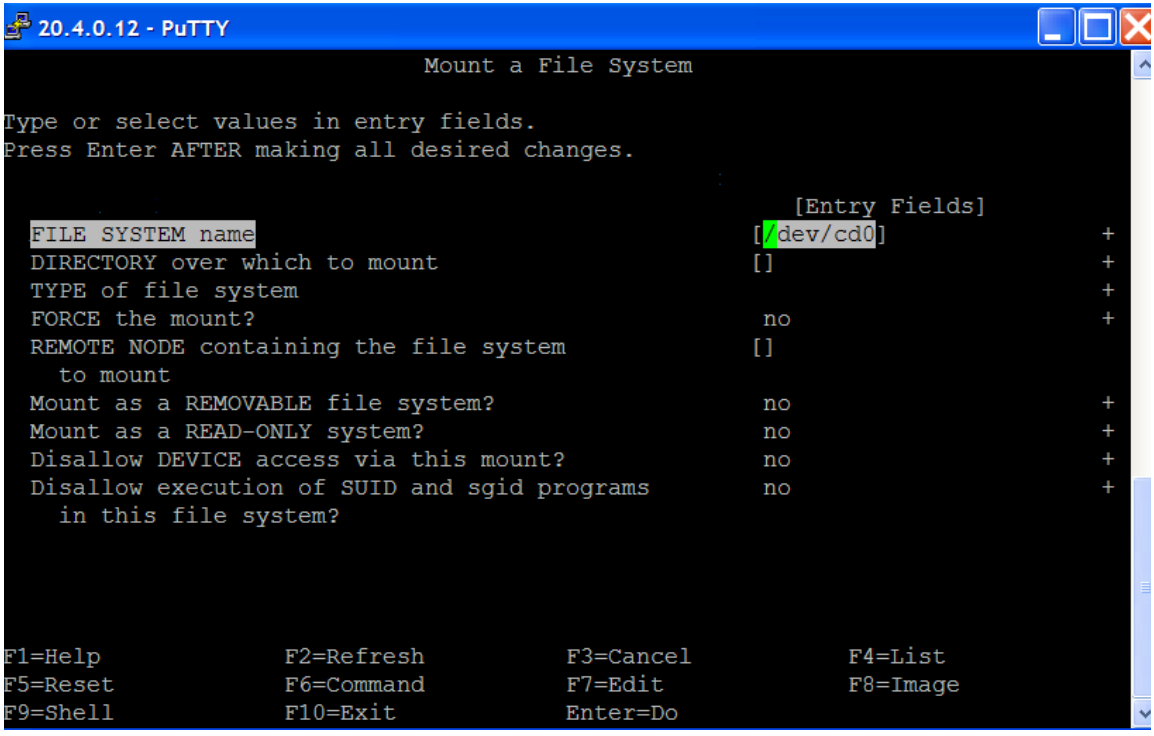
7. Select List All Mounted File Systems.

   The device /dev/cd0 should be mounted.

8. Click Done.

   If the CD-ROM drive is not mounted, you can mount it by selecting Mount A File System, and then selecting /dev/cd0 from the list.

Using SMIT to mount the CD-ROM:


20.4.0.12 - PuTTY
Mount a File System
Type or select values in entry fields.
Press Enter AFTER making all desired changes.
[Entry Fields]
FILE SYSTEM name [/dev/cd0]
DIRECTORY over which to mount []
TYPE of file system
FORCE the mount? no
REMOTE NODE containing the file system []
to mount
Mount as a REMOVABLE file system? no
Mount as a READ-ONLY system? no
Disallow DEVICE access via this mount? no
Disallow execution of SUID and sgid programs in this file system? no
F1=Help F2=Refresh F3=Cancel F4=List
F5=Reset F6=Command F7=Edit F8=Image
F9=Shell F10=Exit Enter=Do

To remove the disk you will need to unmount the CD using SMIT, or type:

```
unmount /usr/cdrom
```

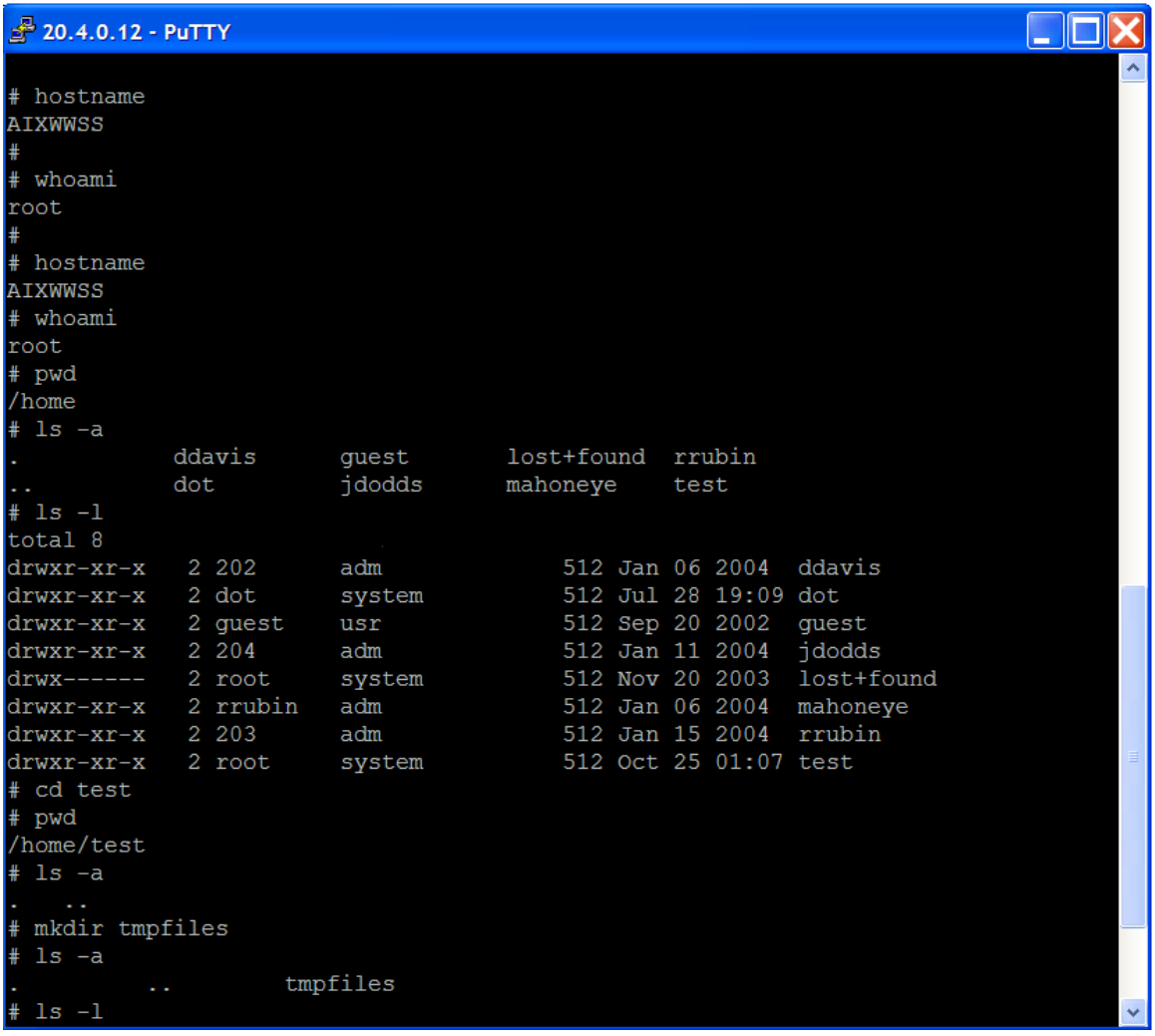
Using navigation and installation commands

There are various commands that assist navigation and installation:

Commands to aid navigation and installation:

Command	Description
hostname	Echoes the system name.
whoami	Echoes the current user name.
pwd	Echoes the current path location.
ls –a	Lists the contents of a directory.
cp	Copies a file or files to another name or location.
cd	Enables the user to change directory, or example cd /tmp changes from the current location to the tmp directory.
mkdir	Creates a new directory in the current location.
chmod	Changes the permissions of a file.
more	Lists the contents of a file.
pg	Lists the contents of a file.

Commands for navigation and installation:



Creating and managing user accounts and groups

Use SMIT to create and manage user accounts and groups. Setting the HOME Directory is very important. A user's HOME Directory should never be the root (/) directory.

The sequence of directories that commands search can be set for all users or for selected users. For all users, it should be included in the /etc/environment file and for selected users it should included in the user's $HOME/.profile file. Because the *.profile file is hidden, use the *ls -a* command to list it.

Use the VI editor to edit these files. It is advised to always make a copy of a file before editing. For example, cp environment environment.original.

E Databanks – Formats

This section describes the various formats that can be utilized when specifying URLs and using databanks in PLGS.

Contents:

Topic	Page
URL addresses	E-2
SPTREMBL flat file format	E-3
Genbank flat file format	E-6
BLAST flat file format	E-8
FASTA flat file format	E-9

URL addresses

The URL address (Uniform Resource Locator) format consists of a Protocol and an Address. Examples of possible protocols are http, ftp, and file. To form the URL, the address is concatenated onto the protocol name, as shown in these examples:

- http://www.someAddress.org/filename.zip
- ftp://www.someOtherAddress.org/directory/flatfile.gz
- file://C:/Directory/subdirectory/sequences.fas

Note that URLs are case sensitive.

SPTREMBL flat file format

The SPTREMBL format is used by Swiss Prot and EMBL.

Example:

```
ID   AI304266   standard; RNA; EST; 187 BP.
XX
AC   AI304266;
XX
SV   AI304266.1
XX
DT   03-JUN-1999 (Rel. 59, Created)
DT   03-JUN-1999 (Rel. 59, Last updated, Version 1)
XX
DE   IpTR040u Channel catfish pituitary library Ictalurus
punctatus cDNA clone
DE   IpTR040 3', mRNA sequence.
XX
KW   EST.
XX
OS   Ictalurus punctatus (channel catfish)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Siluriformes;
OC   Ictaluridae; Ictalurus.
XX
RN   [1]
RP   1-187
RA   Liu Z., Tan G., Li P., Dunham R.;
RT   "Transcribed dinucleotide microsatellites and their
associated genes from
RT   channel catfish, Ictalurus punctatus";
RL   Unpublished.
```

```
XX
DR   UNILIB; 1529; 1529.
XX
CC   Other_ESTs: IpTR040r
CC   Contact: Liu, Z.J.
CC   Fish Molecular Genetics and Biotechnology
CC   Auburn University
CC   203 Swingle Hall, Department of Fisheries, Auburn, AL
36849, USA
CC   Tel: 334 844 4054
CC   Fax: 334 844 9208
CC   Email: zliu@acesag.auburn.edu
CC   Seq primer: M13 forword
CC   High quality sequence stop: 187.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..187
FT                   /db_xref="taxon:7998"
FT                   /db_xref="UNILIB:1529"
FT                   /sex="female"
FT                   /organism="Ictalurus punctatus"
FT                   /strain="Kansas"
FT                   /clone="IpTR040"
FT                   /clone_lib="Channel catfish pituitary
library"
FT                   /tissue_type="pituitary"
FT                   /dev_stage="adult"
XX
SQ   Sequence 187 BP; 58 A; 36 C; 50 G; 43 T; 0 other;
     gggggaaaaa aaccaaacaa acaattacag caggcgcgaa gcaccgatat
cggattagtg        60
     cgtgaacgat accttgagct agtcggtggg acagtcggct aatgctagct
ttgcgattaa       120
```

```
     cgtgtcattc cgagcaagtc ggagcactaa agcagtttgg caaatttaaa
tatgcagttt       180

gagcttt
            187
//
```

Genbank flat file format

The Genbank format is specified by NCBI.

Example:

```
LOCUS       AAC71934                 101 aa
linear   INV 16-APR-2002
DEFINITION  metal binding protein (DHHC domain) [Plasmodium
falciparum 3D7].
ACCESSION   AAC71934
VERSION     AAC71934.1  GI:3845261
DBSOURCE    accession AE001414.1
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7.

ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida;
Plasmodium.
REFERENCE   1  (residues 1 to 101)
 AUTHORS   Gardner,M.J., Tettelin,H., Carucci,D.J.,
Cummings,L.M., Aravind,L.,
            Koonin,E.V., Shallom,S., Mason,T., Yu,K.,
Fujii,C., Pederson,J.,
            Shen,K., Jing,J., Aston,C., Lai,Z.,
Schwartz,D.C., Pertea,M.,
            Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R.,
White,O.,
            Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C.
and Hoffman,S.L.
 TITLE     Chromosome 2 sequence of the human malaria
parasite Plasmodium
            falciparum
 JOURNAL   Science 282 (5391), 1126-1132 (1998)
 MEDLINE   99021743
   PUBMED   9804551
```

```
 REMARK    Erratum:[[published erratum appears in Science
1998 Dec
            4;282(5395):1827]]
REFERENCE   2  (residues 1 to 101)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1998) The Institute for
Genomic Research, 9712
            Medical Center Drive, Rockville, MD 20814, USA
COMMENT     Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..101
                     /organism="Plasmodium falciparum 3D7"
                     /strain="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="2"
     Protein         1..101
                     /product="metal binding protein (DHHC
domain)"
     CDS             1..101
                     /gene="PFB0725c"

/coded_by="complement(join(AE001414.1:1256..1365,

AE001414.1:1500..1634,AE001414.1:1821..1881))"
                     /note="identified by sequence
similarity; putative"
ORIGIN
        1 miiwchikcl ctnpgflnet fhfvsdntte ydnnvqmckk
cnllkikrsh hcsvcdkcim
       61 kmdhhcfwin scvglynqky fillnfvrtk gkyntniikh l
//
```

BLAST flat file format

This format is the same as the NCBI_EXPASY_STANDARD format subtype of FASTA format.

Example:

```
>gi|3845261|gb|AAC71934.1| metal binding protein (DHHC
domain) [Plasmodium falciparum 3D7]
MIIWCHIKCLCTNPGFLNETFHFVSDNTTEYDNNVQMCKKCNLLKIKRSHHCSVCDKCIM
KMDHHCFWIN
SCVGLYNQKYFILLNFVRTKGKYNTNIIKHL
```

FASTA flat file format

FASTA format consists of a description line, beginning with a `>' symbol, followed by multiple lines containing the sequence of amino acid or nucleotide characters.

Example:

```
>gi|3845261|gb|AAC71934.1| metal binding protein (DHHC
domain) [Plasmodium falciparum 3D7]
MIIWCHIKCLCTNPGFLNETFHFVSDNTTEYDNNVQMCKKCNLLKIKRSHHCSVCDKCIM
KMDHHCFWIN
SCVGLYNQKYFILLNFVRTKGKYNTNIIKHL
```

Within this general format, many different conventions are used. If FASTA format is specified as a Databank option, you must also specify the correct FASTA format subtype.

FASTA STANDARD

Description line:

```
>NAME|ACCESSION_NUMBER|DATABANK_OF_ORIGIN: DESCRIPTION
```

Example:

```
>IF3_AQUAE|O67653|SPT: Translation initiation factor IF-3.
MSKLKEYRVNRQIRAKECRLIDENGQQIGIVPIEEALKIAEEKGLDLVEIAPQAKPPVCK
IMDYGKFKYELKKKEREARKKQREHQIEVKDIRMKVRIDEHDLQVKLKHMREFLEEGDKV
KVWLRFRGRENIYPELGKKLAERIINELSDIAEVEVQPKKEGNFMIFVLAPKRKK
```

FASTA NCBI_EXPASY_STANDARD

This format comes in two different forms: a 2-pipe version, and the 4-pipe version shown below. The description line of this particular databank format is not shortened in any way.

Description line:

```
>gi|NUMBER|DATABANK_OF_ORIGIN|ACCESSION_NUMBER|LOCUS_OR_NAME
DESCRIPTION
```

Example of 4-pipe version:

```
>gi|3845261|gb|AAC71934.1| metal binding protein (DHHC
domain) [Plasmodium falciparum 3D7]
MIIWCHIKCLCTNPGFLNETFHFVSDNTTEYDNNVQMCKKCNLLKIKRSHHCSVCDKCIM
KMDHHCFWIN
SCVGLYNQKYFILLNFVRTKGKYNTNIIKHL
```

Example of 2-pipe version:

```
>SP|PLASM_FALCI|(P08978) metal binding protein (DHHC domain)
[Plasmodium falciparum 3D7]
MIIWCHIKCLCTNPGFLNETFHFVSDNTTEYDNNVQMCKKCNLLKIKRSHHCSVCDKCIM
KMDHHCFWIN
SCVGLYNQKYFILLNFVRTKGKYNTNIIKHL
```

FASTA NCBI_PRF_PIR

Description line:

```
>DATABANK_OF_ORIGIN||NAME
```

FASTA NCBI_PDB

Description line:

```
>PDB|NAME|CHAIN
```

Example:

```
>pdb|1IOD|A Chain A, Crystal Structure Of The Complex
Between The Coagulation Factor X Binding Protein From Snake
Venom And The Gla Domain Of Factor X
DCSSGWSSYEGHCYKVFKQSKTWADAESFCTKQVNGGHLVSIESSGEADFVGQLIAQKIK
SAKIHVWIGLRAQNKEKQCS
IEWSDGSSISYENWIEEESKKCLGVHIETGFHKWENFYCEQQDPFVCEA
```

FASTA NCBI_PATENT

Description line:

```
>pat|COUNTRY|NUMBER
```

Example:

```
>pat|US|4772557VAAHELGXSLGLS
```

FASTA NCBI_GENINFO

Description line:

```
>bbs|NUMBER
```

FASTA NCBI_GENERAL

Description line:

```
>gnl|DATABANK_OF_ORIGIN|IDENTIFIER
```

Example:

```
>gnl|spt|O67653 Translation initiation factor IF-3.
MSKLKEYRVNRQIRAKECRLIDENGQQIGIVPIEEALKIAEEKGLDLVEIAPQAKPPVCK
IMDYGKFKYELKKKEREARKKQREHQIEVKDIRMKVRIDEHDLQVKLKHMREFLEEGDKV
KVWLRFRGRENIYPELGKKLAERIINELSDIAEVEVQPKKEGNFMIFVLAPKRKK
```

FASTA NCBI_LOCAL

Description line:

```
>lcl|IDENTIFIER
```

Example:

```
>lcl|O67653 Translation initiation factor IF-3.
MSKLKEYRVNRQIRAKECRLIDENGQQIGIVPIEEALKIAEEKGLDLVEIAPQAKPPVCK
IMDYGKFKYELKKKEREARKKQREHQIEVKDIRMKVRIDEHDLQVKLKHMREFLEEGDKV
KVWLRFRGRENIYPELGKKLAERIINELSDIAEVEVQPKKEGNFMIFVLAPKRK
```

FASTA PDB

Description line:

```
>NAME:CHAIN DESCRIPTION
```

Example:

```
>1C8F:A FELINE PANLEUKOPENIA VIRUS CAPSID
GVGISTGTFNNQTEFKFLENGWVEITANSSRLVHLNMPESENYKRVVVNNMDKTAVKGNM
ALDDIHVEIVTPWSLVDANAWGVWFNPGDWQLIVNTMSELHLVSFEQEIFNVVLKTVSES
ATQPPTKVYNNDLTASLMVALDSNNTMPFTPAAMRSETLGFYPWKPTIPTPWRYYFQWDR
TLIPSHTGTSGTPTNVYHGTDPDDVQFYTIENSVPVHLLRTGDEFATGTFFFDCKPCRLT
HTWQTNRALGLPPFLNSLPQSEGATNFGDIGVQQDKRRGVTQMGNTDYITEATIMRPAEV
GYSAPYYSFEASTQGPFKTPIAAGRGGAQTDENQAADGDPRYAFGRQHGQKTTTTGETPE
RFTYIAHQDTGRYPEGDWIQNINFNLPVTNDNVLLPTDPIGGKTGINYTNIFNTYGPLTA
LNNVPPVYPNGQIWDKEFDTDLKPRLHINAPFVCQNNCPGQLFVKVAPNLTNQYDPDASA
NMSRIVTYSDFWWKGKLVFKAKLRASHTWNPIQQMSINVDNQFNYVPNNIGAMKIVYEKS
QLAPRKLY
```

FASTA PIR

Description line:

```
>ACCESSION PIR1 release RELEASE_NUMBER
```

Example:

```
>S52288  PIR2 release 72.04
MPSKKVLQTEHINTTDEAPKTTSVRPRKRKADVAIHLQDPDEEVTEMTRK
KQCASQACWNPDTGYTSPCRRIPTPDEVEEPVAFGSVGFTQYASESIFIT
PTRSTPLPALCWASKDEVWNNLLGKDKLYLRDTRVMERHPNLQPKMRAIL
LDWLMEVCEVYKLHRETFYLGQDYFDRFMATQENVLKTTLQLIGISCLFI
AAKMEEIYPPKVHQFAYVTDGACTEDDILSMEIIIMKELNWSLSPLTPVA
WLNIYMQMAYLKETAEVLTAQYPQATFVQIAELLDLCILDVRSLEFSYSL
LAASALFHFSSLELVIKVSGLKWCDLEECVRWMVPFAMSIREAGSSALKT
FKGIAADDMHNIQTHVPYLEWLGKVHSYQLVDIESSQRSPVPTGVLTPPP
SSEKPESTIS
```

FASTA SRS

Description line:

>ACCESSION

Example:

>AA917165
cttctagttaaggactgtagaataagcacgcaatataatagagagtacgtgggttttata
atttaattgttcgaatacgttctggatattatcatacttcttcgttcgttcgttatttct
ttcaaaagagttgtaatgaactaaaaacgtataagcaatattcaacttaacaacacaaaa
aag

FASTA ARABIDOPSIS_GENOME

Description line:

>ACCESSION? ENTRY NAME? DESCRIPTION?

Example:

>AT1G69120 68300.M06877 F4N2.9 HOMEOTIC PROTEIN BOI1AP1, PUTATIVE SIMILAR TO HOMEOTIC PROTEIN BOI1AP1 GI:1561777 FROM [BRASSICA OLERACEA]; SUPPORTED BY FULL-LENGTH CDNA: CERES: 39890.
ATGGGAAGGGGTAGGGTTCAATTGAAGAGGATAGAGAACAAGATCAATAGACAAGTGACA
TTCTCGAAAAGAAGAGCTGGTCTTTTGAAGAAAGCTCATG
AGATCTCTGTTCTCTGTGATGCTGAAGTTGCTCTTGTTGTCTTCTCCCATAAGGGAAAAC
TCTTCGAATACTCCACTGATTCTTGTATGGAGAAGATACT
TGAACGCTATGAGAGGTACTCTTACGCCGAAAGACAGCTTATTGCACCTGAGTCCGACGT
CAATACAAACTGGTCGATGGAGTATAACAGGCTTAAGGCT
AAGATTGAGCTTTTGGAGAGAAACCAGAGGCATTATCTTGGGGAAGACTTGCAAGCAATG
AGCCCTAAAGAGCTTCAGAATCTGGAGCAGCAGCTTGACA
CTGCTCTTAAGCACATCCGCACTAGAAAAAACCAACTTATGTACGAGTCCATCAATGAGC
TCCAAAAAAAGGAGAAGGCCATACAGGAGCAAAACAGCAT
GCTTTCTAAACAGATCAAGGAGAGGGAAAAAATTCTTAGGGCTCAACAGGAGCAGTGGGA
TCAGCAGAACCAAGGCCACAATATGCCTCCCCCTCTGCCA
CCGCAGCAGCACCAAATCCAGCATCCTTACATGCTCTCTCATCAGCCATCTCCTTTTCTC
AACATGGGTGGTCTGTATCAAGAAGATGATCCTATGGCAA

```
TGAGGAGGAATGATCTCGAACTGACTCTTGAACCCGTTTACAACTGCAACCTTGGCTGCT
TCGCCGCATGA
```

FASTA NRDB

NRDB is the same subtype as NCBI_EXPASY_STANDARD.

FASTA UNIGENE

Description line:

```
>gnl|UG|UGAccession DESCRIPTION /gb= /gi= /ug= /len=
```

Example:

```
> 2386477 gnl|UG|Hs#S2386477 PM3-FT0024-240500-001-f10 Homo
sapiens cDNA /gb=BE769099 /gi=10222757 /ug=Hs.1287 /len=384
CTCTGAGATCCCCACTTCCAGAGTAGTATAAGATGTTATCCGCCCTCCAGGAGCTTACAA
AACTAGAGGCAGAAATAAGATGTACATGTGACTCAGGCAGCATGTGACACACACAAAGGT
GGGCAGCTCTGAGACAATGGTGGTCAAGTGACCACTGAGGCCCAGAGCCGTTGGAACAGT
CTCTTAGAACAGGGTGGAGGACTTAAAACTTGGATGAACAGGGGCTGGCAGAGCACTTGG
AATGGGTAAGGACAAGACCGGGAGATCAATTTGGCTGGAGCAGGGGAGCTTGTGTTATAT
ATGCAGAAAAAGGTTGAAACGGGGAAGTTTTAATACTGTTTAGGTAAATAAGGATTAAAC
ACAAAAGGAAGGAAAAACGTGAGA
```

FASTA STANDARD_SPACED

Description line:

```
>NAME ACCESSION_NUMBER DESCRIPTION
```

Example:

```
>IF3_AQUAE (O67653) Translation initiation factor IF-3.
MSKLKEYRVNRQIRAKECRLIDENGQQIGIVPIEEALKIAEEKGLDLVEIAPQAKPPVCK
IMDYGKFKYELKKKEREARKKQREHQIEVKDIRMKVRIDEHDLQVKLKHMREFLEEGDKV
KVWLRFRGRENIYPELGKKLAERIINELSDIAEVEVQPKKEGNFMIFVLAPKRKK
```

FASTA LONG_DESCRIPTION

Description line:

```
>NAME DESCRIPTION
```

This format is used when the description is very long. In the ProteinLynx display, the description is truncated to fit into the viewing area.

Example:

```
>gp:AL034396_1 PID:5441319 Human DNA sequence from clone
1158B12 on chromosome Xp11.21-11.4 Contains the ZXDA gene
for X-linked duplicated Zinc finger A, and MYCL1 (v-myc
avian myelocytomatosis viral oncogene homolog 1, lung
carcinoma derived) and KRT8 (Keratin 8, Cytokeratin 8, CYK8,
Keratin type II skeletal 8) pseudogenes. Contains ESTs, an
STS, GSSs and a CpG island, complete sequence; match:
proteins: Sw:P98168 Sw:P98169. (gb:AL034396)
MEIPKLLPARGTLQGGGGGGIPAGGGRVHRGPDSPAGQVPTRRLLLPRGPQDGGPGRRRE
EASTASRGPGPSLFAPRPHQPSGGGDDFFLVLLDPVGGDVETAGSGQAAGPVLREEAKAG
PGLQGDESGANPAGCSAQGPHCLSAVPTPAPISAPGPAAAFAGTVTIHNQDLLLRFENGV
LTLATPPPHAWEPGAAPAQQPRCLIAPQAGFPQAAHPGDCPELRSDLLLAEPAEPAPAPA
PQEEAEGLAAALGPRGLLGSGPGVVLYLCPEALCGQTFAKKHQLKMHLLTHSSSQGQRPF
KCPLGGCGWTFTTSYKLKRHLQSHDKLRPFGCPAEGCGKSFTTVYNLKAHMKGHEQENSF
KCEVCEESFPTQAKLGAHQRSHFEPERPYQCAFSGCKKTFITVSALFSHNRAHFREQELF
SCSFPGCSKQYDKACRLKIHLRSHTGERPFLCDFDGCGWNFTSMSKLLRHKRKHDDDRRF
MCPVEGCGKSFTRAEHLKGHSITHLGTKPFVCPVAGCCARFSARSSLYIHSKKHLQDVDT
WKSRCPISSCNKLFTSKHSMKTHMVKRHKVGQDLLAQLEAANSLTPSSELTSQRQNDLSD
AEIVSLFSDVPDSTSAALLDTALVNSGILTIDVASVSSTLAGHLPANNNNSVGQAVDPPS
LMATSDPPQSLDTSLFFGTAATGFQQSSLNMDEVSSVSVGPLGSLDSLAMKNSSPEPQAL
TPSSKLTVDTDTLTPSSTLCENSVSELLTPAKAEWSVHPNSDFFGQEGETQFGFPNAAGN
HGSQKERNLITVTGSSFLV
```

FASTA ACCESSION_ONLY

Description line:

```
>ACCESSION
```

Example

```
>AA917165
cttctagttaaggactgtagaataagcacgcaatataatagagagtacgtgggttttata
atttaattgttcgaatacgttctggatattatcatacttcttcgttcgttcgttatttct
ttcaaaagagttgtaatgaactaaaaacgtataagcaatattcaacttaacaacacaaaa
aag
```

Index

Symbols

*.csv 11-2, 11-4, 11-6, 11-9, 11-10
*.dta 2-22
*.gz 13-8
*.html 11-2, 11-4, 11-6, 11-9, 11-10
*.jar 2-26
*.mstext 2-22
*.mzdata 2-23
*.olb 5-30
*.pkl 2-22, 14-5
*.txt 9-4
*.xls 9-4
*.xml 7-10, 8-4, 14-5
*.xsl 7-11
*.Z 13-8
*.z 13-8
*.zip 13-8

A

acquiring data 5-31, A-11, A-21
    Electrospray DDA 8-5
    Electrospray High/Low 8-5
    Electrospray MS 8-5
    MALDI MS 8-5
    MALDI PSD MX 8-5
    MALDI Q-Tof MS 8-5
    MALDI Q-Tof MSMS 8-5
Acquisition tab 16-4
ACTH A-5
Add Bookmark dialog box 2-11
add/remove columns
    peptide or protein table 6-14
Add/Remove Tools dialog box 2-4
adding
    databanks 13-3
    digest reagents 12-9
    export plugins 2-24
    gel spots 9-3
    gels 9-3
    inlet file 16-7
    method file 16-7
    modifier reagents 12-4
    new databank A-25
    processing parameters 5-21
    sample 4-2
    search engines 2-6
    workflow templates 5-7, 5-20
ADH A-5
AIX
    installation 1-15
    starting PLGS 1-19
algorithm
    BLOSUM 14-25
    PAM 14-25
annotating samples 5-11
Applies to attribute 12-5
archived databanks
    restoring 13-15
archives 13-10
    databanks 13-15
    deleting files 13-14
    deleting revived archives 13-14
assess data quality
    Expression experiment 10-8
assess data quality viewer 10-25
associated masses 6-34, 6-35
attaching
    data processing parameters A-8, A-18
    raw data 5-13
    workflow file A-9, A-19
    workflow templates 5-20
attribute sets

Chromatogram 8-5, 8-15
Deisotoping and Centroiding 8-5, A-14
Mass Accuracy 8-5, A-14
Noise Reduction 8-5, A-14
Peak Matching 8-5
attributes
Applies to 12-5
Automatic Thresholds 8-13, 8-16
Background Polynomial 8-10
Background Subtract Type 8-9
Background Threshold 8-9
Calibration File 8-15
Centroid Top 8-13
Combine Options 8-10
databank 13-4
Deisotoping type 8-12
Delta Mass 12-5
Expected Peak Width 8-15
External Lock Mass 8-6
Fragment Intensity Threshold 8-15
Fragment Matching Window 8-15
Fragments 12-5
Intensity Range 8-11
Intensity Threshold 8-7
Iterations 8-12
Lock Mass tolerance 8-7
Lock Spray Lock Mass 8-8
Lock Spray Scans 8-8
Low Mass Threshold 8-11
Maximum Number of Charges 8-14
Minimum Charges to Report 8-14
Minimum Peak Width 8-13, 8-15
Modifier type 12-5
Name 12-4
NP Multiplier 8-14
Number of Precursors 8-15
Peak Width Units 8-16
Peptide Filter 8-11
Perform Deisotoping 8-12
Perform Lock Spray Calibration 8-8
Perform Smoothing 8-10
Precursor Matching Window 8-15
Primary Internal Lock Mass 8-7
Quantitation Reagent 12-5
Range Units 8-16
Report Monoisotopic Fragments 8-15
Scans to Combine 8-10
Secondary Internal Lock Mass 8-7
Select Calibration Type 8-6
Select start time 8-16
Select stop time 8-16
Select time range 8-16
Smoothing Iterations 8-10
Smoothing Type 8-10
Smoothing Window 8-10
Threshold 8-13, 8-16
Threshold Type 8-7
TOF Resolution 8-14
automated task
AutoMod Query 7-9
BLAST Query 7-9
Databank Search 7-8
De Novo Query 7-9
automatic data curation 6-12, B-7
Automatic Thresholds attribute 8-13, 8-16
Automation Setup dialog box 2-18, 2-24
AutoMod Analysis 14-14–14-18
Consider Modifications parameter 14-16
Consider Substitutions parameter 14-16
search parameters 14-16
validate results 14-17

AutoMod Analysis search parameters 14-15
AutoMod Analysis tool 14-1, 14-14
AutoMod Query
automated task 7-9
filter 7-11
average 14-12
axis
assess data quality 10-25

B

Backed-up folder
restoring 1-5, 1-11
Background Polynomial attribute 8-10
background subtract type 8-9
Background Subtract Type attribute 8-9
Background Threshold attribute 8-9
backing up
PLGS folders in Linux 1-7
PLGS folders in Windows 1-3
BLAST 6-5
BLAST Query 6-7
BLAST View 6-7
make blastable 13-6, 13-10
results 6-7, 14-26
results panel 14-27
BLAST algorithm
search parameters 14-24
BLAST flat file format E-8
BLAST Searching tool 13-6, 14-1, 14-23–14-27
blastable 13-6, 13-10
blocking mode 2-20
BLOSUM
algorithm 14-25
matrices 14-25
bookmarks
modifying 2-12
removing 2-12
buttons
Delete 4-2, 7-4, 12-6, 12-10, 13-13, 13-14, 13-15
Remove 2-8, 2-9, 2-12, 7-4, 8-4, 13-13
Save 5-22, 7-4, 8-3, 12-5, 12-10, 13-12

C

Calibration File attribute 8-15
calibration type
select 8-6
centroid top 8-13
Centroid Top attribute 8-13
change column order 6-15
peptide or protein table 6-14
Change Processing Parameters command 5-21
changing
preferences 2-5
processing parameters 5-7
Chromatogram attribute set 8-5, 8-15
circled gel spots 9-9
Clear OK assignments 6-5
client
installation 1-3
starting PLGS 1-5
client/server environment, installation 1-1
closing projects 3-6
clusters
import significant 10-24
include or exclude 10-13
Coarse Delta retention time 5-27
columns
displaying 6-14
Combine Options attribute 8-10
commands

Change Processing Parameters 5-21
Import Worksheet 5-30, A-10, A-20
Microkernel Search Engine 15-15
Process Raw Data 5-17
Compression Type 13-8, 13-9
confidence limit filter 10-15
connecting
to search engine 13-17
Consider Modifications parameter 14-16
Consider Substitutions parameter 14-16
Container Manager 5-2
copying data 6-16, 6-26
creating
databanks 13-3
Expression experiment 10-3
new project 3-2, A-2
new target plate 5-9
project 5-3, 10-2
target plate 5-9
workflows A-7, A-17
cross
OK column 6-12
curated filter
print templates 11-16
curation
automatic 6-12, B-7
data 6-5
of data 10-11

D

data
acquisition 15-1, A-11, A-21
curation 10-11
automatic 6-12, B-7
Expression 16-1
file 14-5
graphical 11-14
MS^E 16-1
printing 11-2
processing 15-14
tabular 11-14
data directed analysis (DDA) A-23
chromatograms A-22
Data Preparation tool
attribute sets 8-5
creating a new processing parameters template 8-2
definition of screen areas 8-3
processing parameters templates 8-5
removing processing parameters templates 8-4
saving processing parameters templates 8-3
select data type 8-2
data quality viewer 10-25
data type
MS 7-2
MSMS 7-2
PSD 7-2
Databank 14-1
databank
archives
reviving 13-15
attributes 13-4
Databank Admin tool 13-2, 13-2–13-17
description 13-2
databank attribute
Download Compression Type 13-8
Download Renew Period 13-8
Download URL Address 13-8
FASTA Format 13-5
Format 13-4
Index For PepGrab 13-6
Keep Archives 13-10

Load into Memory 13-6
Location 13-6
Make Blastable 13-6
Management Options 13-7
Name 13-4
Periodically Download 13-7
Periodically Update 13-9
Processing End Time 13-10
Processing Start Time 13-10
Species for Indexing 13-7
Type 13-4
Update Compression Type 13-9
Update Renew Period 13-10
Update URL Address 13-9
Databank Search 14-3–14-13
automated task 7-8
parameters 14-5
tool 14-3–14-13
Databank Search parameters 14-5
Data File 14-5
Databanks 14-6
Database 14-6
Enzyme 14-9
Estimated Calibration Error 14-7
Exclude Masses 14-11
Fixed Modifications 14-10
Fragment Tolerance 14-7
Instrument Type 14-13
Mass Spectrum 14-5
Mass Values 14-12
Maximum Hits to Return 14-9
Minimum Peptides to Match 14-9
Missed Cleavages 14-10
Molecular Weight Range 14-8
Monoisotopic or Average 14-12
MSMS Tolerance 14-7
Peptide Charge 14-12
Peptide Tolerance 14-6
pI Range 14-8
PLGS 14-4
Primary Digest Reagent 14-9
Protein Mass 14-8
Search Engine Type 14-5
Secondary Digest Reagent 14-10
Species 14-6
Taxonomy 14-6
Validate Results 14-12
Variable Modifications 14-11
databank searching
real time 15-1
databanks 14-6
adding 13-3
archives 13-15
creating 13-3
deleting 13-13
editing 13-11
hyperlinks 4-4
real time searching 15-1
removing 13-13
restoring old 1-23
retrieving entries 6-30
search 14-3–14-13
database 14-6
data-dependent acquisition. See DDA
DDA 15-1, 15-8, A-22, A-23
DDA file
setting up 15-10
De Novo Query 14-19–??
automated task 7-9
filter 7-11
sequencing parameters 14-21
De Novo Sequencing
parameters 14-20
tool 14-1, 14-19
validate results 14-22
deisotope
peak detection 15-11
type 8-12

Deisotoping and Centroiding attribute set 8-5
Deisotoping type attribute 8-12
Delete button 4-2, 7-4, 12-6, 12-10, 13-13, 13-14, 13-15
deleting
archive files 13-14
databanks 13-13
digest reagents 12-10
modifier reagents 12-6
print templates 11-12
projects 3-6
sample 4-2
Delta Mass attribute 12-5
descriptions
Databank Admin tool 13-2
Digest Reagent tool 12-7
Expression Analysis tool 10-2
Gel Manager 9-2
Print tool 11-2
processing parameters templates 8-2
Sample Manager tool 4-2
Design Manager
Expression analysis 10-3
diagnostics
displaying
real time 15-15
showing 15-15
windows 15-15
dialog boxes
Add Bookmark 2-11
Add/Remove Tools 2-4
Automation Setup 2-18, 2-24
Import Gel Spots 9-3
Installation Successful 1-5
Modify Bookmark 2-12
Modify Processor 2-9
Modify Search Engine 2-7
New Container Tool 5-9
PeptideAuto Server 5-31
PlugIn Selector 2-25
ProteinLynx Browser Automation Setup 2-18
ProteinLynx Browser Preferences 2-5, 5-33
Select a Colour 2-14, 2-15
Select Files 5-15
single 5-14
Select Processing Parameters A-9, A-19
Specify Plates 9-4
Start Sample List Run 5-31, A-11, A-21
URL Chooser 7-10
digest fragments
Protein Workpad 6-30
Digest Reagent tool 12-7–12-10
description 12-7
digest reagents
adding 12-9
deleting 12-10
editing 12-9
non-specific 14-10
saving 12-10
viewing 12-8
displaying
columns 6-14
ion probabilities 6-22
real time diagnostics 15-15
displays
PeptideAuto Server A-12
docs folder 1-5, 1-11
Download Compression Type databank attribute 13-8
Download Renew Period databank attribute 13-8

Download URL Address databank attribute 13-8
downregulation 10-11
dta format 2-22
dynamic bookmark 2-11

E

edit precision
  peptide or protein table 6-14
editing
  databanks 13-11
  digest reagents 12-9
  modifier reagents 12-4
  workflow templates 7-9
Electrospray DDA 8-5
Electrospray High/Low 8-5
Electrospray MS 7-2, 7-5, 8-5
Electrospray Shotgun 7-2, 7-5
EMBL E-3
EMRT table 10-9
  export switch lists 10-23
  import significant clusters 10-24
  view replicates for cluster 10-12
  viewing 10-10
End Time 13-10
enzyme 14-9
Error Messages 6-19
erythromycin A-14
EST
  data 6-3
  table 6-12
EST sequences
  selecting for search 14-18
estimated calibration error 14-7, 14-21
E-value 14-26
Excel files (.xls) 9-4
exclude clusters 10-13
exclude masses 14-11
  viewing 6-34
  workpad 6-31
Exclude Masses Workpad 6-31
  Masses to Exclude window 6-34
executable
  file for Windows 1-4
Expect Threshold parameter 14-25
Expected Peak Width attribute 8-15
experiment attributes
  Expression 10-4
experiment setup
  Expression 16-3
  MS^E 16-3
Export PlugIns 2-23
exporting
  data 11-2
  Expression results 10-12
  mass spectra 5-22
  projects 3-3
  sample list 5-29, A-9, A-19
  spectra 5-28
  SuperTrack results 5-28
  switch lists 10-23
Expression
  assess data quality 10-25
  data 7-2, 7-5, 16-1
  exporting results 10-12
  filtering results 10-13
  method file 16-3
  printing results 10-13
Expression Analysis Design Manager 10-3
Expression Analysis tool 10-2
  creating a project 10-2
  description 10-2
Expression experiment
  assess data quality 10-8
  attributes 10-4
  manually assign samples to groups 10-7

manually define experiment variables 10-6
new 10-3
open 10-3
quantitation analysis 10-8
select data 10-7
select grouping method 10-5
starting 10-9
viewing results 10-10
Expression tab 16-5
Expression table
opening 10-10
external lock mass 8-6
External Lock Mass attribute 8-6

F

FASTA flat file format E-9
FASTA format 14-18
FASTA Format databank attribute 13-5
file format
significant clusters 10-24
file formats
dta 2-22
mass spectrum 14-5
mstext 2-22
mzData 2-23
PDQuest 9-4
pkl 2-22
PKL, mass spectrum 14-5
XML, workflow templates 7-10
XSL 7-11
file permissions
changing 1-8
filter
Expression results 10-13
confidence limit 10-15
P value 10-15
ratio 10-15
replicate 10-14
upregulation 10-15
filters 7-11, 14-26
De Novo Query 7-11
for workflow 7-11
print templates
curated 11-16
numeric 11-16
text 11-16
XML 7-11
Fine Delta retention time 5-27
fixed modifications 14-10, 14-16
format
FASTA 14-18
significant clusters file 10-24
Format databank attribute 13-4
fragment
ion display 6-20
tolerance 14-7, 14-16, 14-21
fragment data
low and high energy 7-5
Fragment Intensity Threshold attribute 8-15
Fragment Matching Window attribute 8-15
Fragments attribute 12-5

G

gapped 14-26
gel
adding 9-3

image 9-6
location of gel spots 9-9
manipulating 9-9
showing axis labels 9-9
viewing 9-9
zooming 9-9
importing 9-3
importing from PDQuest XML file 9-6
importing from Progenesis XML file 9-6
results
viewing 9-9
spots
adding without image 9-3
circled 9-9
importing 9-3
location on gel image 9-9
Gel Manager 4-5, 9-2–9-10
description 9-2
processing data 9-8
replacing a sample 9-7
Genbank flat file format E-6
generating processed samples 4-5
glu-fibrinopeptide B A-14
Graphical Data 11-14

H

high energy fragment data 7-5
homology threshold B-7
host 2-20
hyperlinks
databanks 4-4

I

ICAT experiments 10-21
icons
AutoMod Analysis 14-14
BLAST Searching 14-23
Container Manager 5-2
Data Preparation tool 8-2
Databank Search 14-3
Databank Searching 15-5
Digest Reagent 12-7
real time status 15-7
sample list view column 5-7
spectrum 5-18
workflow 5-18
WorkFlow Designer 7-2
identity threshold B-7
Import Gel Spots dialog box 9-3
Import Mass Spectrum parameter 5-24
Import PlugIns 2-23
Import Worksheet command 5-30, A-10, A-20
importing
gel 9-3
gel spots 9-3
mass spectra 5-22
projects 3-3
significant clusters 10-24
include clusters 10-13
index for PepGrab 13-6
Index For PepGrab databank attribute 13-6
influence 6-23
Installation troubleshooting on UNIX 1-20
installing
in a client/server environment 1-1
on AIX 1-15
on Linux 1-7
on Windows 1-3
services 1-4
instrument
specifications A-1, B-1
type 14-13
Intensity Range attribute 8-11

Intensity Threshold attribute 8-7
interfacing with MassLynx 5-29
internal standards 10-9, 10-10
ion
  display fragment 6-20
  probabilities 6-22
IP address 1-4, 2-6, 2-20
isobaric experiments 10-21
isotope-labeled samples 10-5
iterations 8-12
Iterations attribute 8-12
iTRAQ experiments 10-21

K

Keep Archives databank attribute 13-10

L

label-free analysis 10-5
Link from BLAST Results parameter 2-12
Linux
  installation 1-7
Load into Memory databank attribute 13-6
Location databank attribute 13-6
lock mass
  external 8-6
  lockspray 8-8
  primary internal 8-7
  secondary internal 8-7
  tolerance 8-7
Lock Mass tolerance attribute 8-7
Lock Spray Lock Mass attribute 8-8
Lock Spray Scans attribute 8-8
LockMass tab 16-6
lockspray
  lock mass 8-8
Log files
  Linux 1-13
  UNIX 1-19
  Windows 1-6
low energy fragment data 7-5
Low Mass Threshold attribute 8-11

M

Make Blastable databank attribute 13-6
MALDI
  scoring B-4
  test procedure A-5
MALDI MS 8-5
MALDI PSD MX 8-5
MALDI Q-Tof MS 8-5
MALDI Q-Tof MSMS 8-5
  processing parameters templates 8-5
Management Options databank attribute 13-7
manually assign samples to groups
  Expression experiment 10-7
manually define experiment variables
  Expression experiment 10-6
Manually starting modules
  on Linux 1-13
  on UNIX 1-19
  on Windows 1-6
Mascot
  results 6-5
  search engine 7-6
  simplifying peaks for 5-26
mass
  error 6-23
  spectrum 14-16, 14-21
Mass Accuracy attribute set 8-5
mass spectra 14-5
  exporting 5-22
  importing 5-22
  viewing processed 5-19

mass values 14-12
masses monoisotopic 5-19
Masses to Exclude window 6-34
masses view 6-5, 6-7
MassLynx 5-29
  Acquisition 15-9
  sample list 5-29, A-11
MassLynx Directory parameter 2-19
matrices
  BLOSUM 14-25
  PAM 14-25
  scoring 14-25
MaxEnt
  Lite 15-5
    parameter 15-5
maximum
  hits to return 14-9
  substitutions 14-16
Maximum Number of Charges
    attribute 8-14
Mean Smoothing 8-10
Merge Results parameter 5-23
merging
  MSMS Spectra 5-24
method file
  Expression 16-3
  MS^E 16-3
Microkernel Search Engine command
    15-15
Minimum Charges to Report attribute
    8-14
Minimum Peak Width attribute 8-13,
    8-15
minimum peptides to match 14-9
missed cleavages 14-10, 14-16
modifications to peptides
  specifying 14-21
modifier reagents
  adding 12-4
  deleting 12-6
  saving 12-5
  viewing 12-3
Modifier Tool 12-2–12-6
Modifier type attribute 12-5
Modify Bookmark dialog box 2-12
Modify Processor dialog box 2-9
Modify Search Engine dialog box 2-7
modifying
  bookmarks 2-12
  processors 2-9
  sample 4-3
  search engines 2-7
Modules
  starting manually on Linux 1-13
  starting manually on UNIX 1-19
  starting manually on Windows 1-6
molecular weight range 14-8
monoisotopic 14-12
  masses 5-19
MS Data A-10, A-20
MS Method 5-30, A-5, A-14
MS Method Editor 15-3
MS Text format 2-22
MS^E
  data 7-2, 7-5, 16-1
  function 16-2
  method file 16-3
MSMS
  tolerance 14-7
multiple
  associated masses 6-35
  fixed modifications 14-11
  species 14-6
  variable modifications 14-11
mzData format 2-23

N

Name attribute 12-4

Name databank attribute 13-4
NanoLockSpray 16-2
navigator tree 6-2, 6-9
  results browser 6-7
NCBI E-6
New Container Tool dialog box 5-9
new databank
  adding A-25
New Expression experiment 10-3
new project
  creating A-2
noise reduction
  Q-Tof MSMS A-16
Noise Reduction attribute set 8-5
non-specific digest reagent 14-10
normalization
  automatic 10-9
  internal standards 10-9
NP Multiplier attribute 8-14
Number of Precursors attribute 8-15
numeric filter
  print templates 11-16

O

OK column
  cross 6-12
  question mark 6-12
  tick 6-12
OK filter 6-5, 10-12
opening
  Expression experiment 10-3
  Expression table 10-10
  print templates 11-12
  projects 3-5
organizing samples 5-11

P

P value filter 10-15
PAM
  algorithm 14-25
  matrices 14-25
parameters
  AutoMod Analysis 14-15
  BLAST algorithm 14-24
  Consider Modifications 14-16
  Consider Substitutions 14-16
  Databank Search
    PLGS 14-4
  De Novo Sequencing 14-20, 14-21
  Expect Threshold 14-25
  FASTA Format 13-5
  Import Mass Spectrum 5-24
  Link from BLAST Results 2-12
  MassLynx Directory 2-19
  MaxEnt Lite 15-5
  Merge Results 5-23
  Peak Centering 15-5
  PeptideAuto Port 2-19
  Process Method 15-5
  Smooth 15-5
  Subtract 15-5
  View Results 5-23
PDQuest
  files 9-4
  XML 3-3
  XML file
    importing gels from 9-6
Peak Centering parameter 15-5
Peak Matching attribute set 8-5
Peak Width 8-13
Peak Width Units attribute 8-16
peaks
  simplifying 5-26
PepGrab 6-11
PepGrab View 6-11
peptide
  charge 14-12
  data 6-3
  sequence 14-25

table 6-15
tolerance 14-6, 14-16
view 6-5, 6-7, 6-9
Peptide Filter attribute 8-11
peptide table 6-13
add/remove columns 6-14
change column order 6-15
PeptideAuto Port parameter 2-19
PeptideAuto Server dialog box 5-31
PeptideAuto Server display A-12
peptides
specifying modifications 14-21
Perform Deisotoping attribute 8-12
Perform Lock Spray Calibration attribute 8-8
Perform Smoothing attribute 8-10
Periodically Download databank attribute 13-7
Periodically Update databank attribute 13-9
PKL 14-5
format 2-22
pl range 14-8
plain text files (*.txt) 9-4
plate colors
defaults 2-13
Plate View 5-23
PLGS folders
backing up in Linux 1-7
backing up in Windows 1-3
PLGS search engine 7-6
PLmicokernel 15-15
PlugIn Selector dialog box 2-25
PlugIns
Export 2-23
adding 2-24
Import 2-23
replacing 2-24
preferences
changing 2-5
previously acquired data
processing A-2
primary digest reagent 14-9, 14-16, 14-21
Primary Internal Lock Mass attribute 8-7
print templates
curated filter 11-16
deleting 11-12
numeric filter 11-16
opening 11-12
text filter 11-16
Print tool 11-2–11-25
description 11-2
Print Wizard 6-16, 10-13, 11-3
print workflow 6-16
printing 11-2
Expression results 10-13
opening and deleting templates 11-12
project template 11-2
results 6-16
templates 11-2
workflow template 11-2
probability of upregulation filter 10-15
Process Mass Spectrum 5-7
Process Method parameter 15-5
Process Raw Data 5-7
Process Raw Data command 5-17
process_kernel 15-15
processed
spectrum 5-19
Processed Data Viewer 5-19
processed samples
generating 4-5
processing

data
from a sample list 5-7
Gel Manager 9-8
parameters 15-4
previously acquired data A-2
Processing End Time databank attribute 13-10
processing parameters 5-2, 5-6
adding 5-21
changing 5-7
MALDI
attaching A-8
Q-Tof MSMS
attaching A-18
setting A-14
setting A-6
specifying 5-15
processing parameters templates 5-21, 8-5
attribute sets
Chromatogram 8-5, 8-15
Deisotoping and Centroiding 8-5
Mass Accuracy 8-5
Noise Reduction 8-5
Peak Matching 8-5
creating 8-2
description 8-2
methods to acquire data 8-5
removing 8-4
saving 8-3
Processing Start Time databank attribute 13-10
processors
host 2-20
modifying 2-9
port 2-20
removing 2-9
Progenesis XML file 3-3
importing gels from 9-6
program group 1-5
project template
printing 11-2
projects 3-1
closing 3-6
creating 3-2, 5-3, 10-2
deleting 3-6
exporting 3-3
importing 3-3
opening 3-5
updating 3-5
Protein Expression 10-2
protein mass 14-8
protein sequences
selecting for search 14-18
Protein table 6-12, 10-9, 10-13
add/remove columns 6-14
change column order 6-15
view replicates 10-12
viewing 10-10
Protein view 6-4, 6-7
Protein Workpad 6-27
digest fragments 6-30
ProteinLynx Browser Automation Setup dialog box 2-18
ProteinLynx Browser Preferences dialog box 2-5, 5-33

Q

quantitation
assess data quality 10-25
quantitation analysis
Expression experiment 10-8
Quantitation Reagent attribute 12-5
query tools
description 14-1
toolbars 14-2
question mark 6-5

OK column 6-12

R

Range Units attribute 8-16
ratio filter 10-15
raw data 5-17
attaching 5-13
reagents
modifier 14-21
Real Time
data processing 15-14
databank searching 15-1, 15-8
setting up 15-8
displaying diagnostics 15-15
menu 15-8
status 15-7, 15-9
real time status 15-10
remote searching 15-14
Remove button 2-8, 2-9, 2-12, 7-4, 8-4, 13-13
remove/add columns
peptide or protein table 6-14
removing
bookmarks 2-12
databanks 13-13
processors 2-9
search engines 2-8
Renew Period 13-8, 13-10
replacing Import PlugIns 2-24
replicate filter 10-14
replicates
viewing for a cluster/protein 10-12
Report Monoisotopic Fragments attribute 8-15
required columns
MS^E sample list 16-7
requirements for sample lists 5-4
restoring
archived databanks 13-15
backed-up folder 1-5, 1-11
old databanks 1-23
resubmitting search 6-15
results
browser 6-3
export Expression 10-12
filter Expression 10-13
print Expression 10-13
viewing 6-2
results panel
BLAST 14-27
retrieving databank entries 6-30
reviving databank archives 13-15
root folder 1-5, 1-11
rtdb_monitor 15-15
running
a simulated digest 6-29
MS^E sample list acquisition 16-8
on AIX 1-19
on the server 1-19

S

Sample Editor 4-3
sample lists 5-2, A-11
columns 5-4
custom values 5-5
exporting 5-29
importing 5-3
processing and searching data 5-7
required columns
MS^E acquisition 16-7
requirements 5-4
view column 5-7
viewing 5-5
Sample Manager tool 4-2, 5-11
description 4-2
samples
adding 4-2
deleting 4-2
modifying 4-3

organizing and annotating 5-11
viewing annotation 9-10
viewing information 5-23
Save button 5-22, 7-4, 8-3, 12-5, 12-10, 13-12
saving
digest reagents 12-10
modifier reagents 12-5
Savitzky-Golay 8-10
Scans to Combine attribute 8-10
scoring
MALDI B-4
matrices 14-25
matrix 14-25
schemes B-1
summary B-2
Search Engine
tab 2-5
search engines
adding 2-6
connecting to 13-17
Mascot 7-6
modifying 2-7
PLGS 7-6
removing 2-8
type 14-5
search method
AutoMod Analysis 7-2
BLAST Searching 7-2
Databank Search Query 7-2
De Novo Sequencing 7-2
search parameters
databank 14-5
for BLAST algorithm 14-24
search type
Fragment Ion Search 7-2
PMF (Peptide Mass Fingerprinting) 7-2
PMF + Fragment Ion Search 7-2
searching
methods 7-2
parameters 15-5
strategy 7-2
searching data
from a sample list 5-7
secondary digest reagent 14-10, 14-16, 14-21
secondary internal lock mass 8-7
Secondary Internal Lock Mass attribute 8-7
Select a Colour dialog box 2-14, 2-15
Select Files dialog box 5-14, 5-15
Select Processing Parameters dialog box A-9, A-19
Select start time attribute 8-16
Select stop time attribute 8-16
Select time range attribute 8-16
selecting
data
Expression experiment 10-7
EST 6-15
EST sequences for search 14-18
grouping method
Expression experiment 10-5
peptides 6-15
protein sequences for search 14-18
proteins 6-15
URL 14-5
selecting calibration type 8-6
sequencing De Novo parameters 14-21
server
starting PLGS 1-19
services
installing 1-4
Set Raw Data 5-13, 5-15
setting
processing parameters A-6
samples 5-11

showing
axis labels 9-9
diagnostics 15-15
significant clusters
import 10-24
simulated digest 6-33
running 6-29
Smooth parameter 15-5
Smoothing Iterations attribute 8-10
Smoothing Type attribute 8-10
smoothing types
Mean Smoothing 8-10
Smoothing Window attribute 8-10
species 14-6
Species for Indexing databank attribute 13-7
specifier 12-10
Specify Plates dialog box 9-4
specifying
estimated calibration error 14-21
maximum hits 14-21
maximum substitutions 14-16
processing parameters 5-15
substitutions and modifications per peptide 14-16
templates 5-15
workflow templates 5-15
spectrum
icons 5-18
viewing 5-19
Spectrum Output tab 2-20
Spectrum Viewer 6-3
MS Data 6-16
MSMS Data 6-21
Options 6-24
SPTREMBL flat file format E-3
Start Sample List Run dialog box 5-31, A-11, A-21
Start Time 13-10
starting
Expression experiment 10-9
MassLynx Acquisition 15-9
modules manually
on Linux 1-13
on UNIX 1-19
on Windows 1-6
MS^E sample list acquisition 16-8
PLGS on a client 1-5
PLGS on a single PC 1-6
PLGS on AIX 1-19
static bookmark 2-11
Subtract parameter 15-5
summary scoring B-2
SuperTrack 5-18, 5-26
exporting results 5-28
Swiss Prot E-3
switch lists
export 10-23

T

table
EST 6-12
tabs
Search Engine 2-5
Spectrum Output 2-20
Tabular Data 11-14
target plate
creating new 5-9
taxonomy 14-6
templates
specifying 5-15
test procedure
MALDI A-5
text filter
print templates 11-16
threshold
homology B-7
identity B-7

Threshold attribute 8-13, 8-16
Threshold Type attribute 8-7
tick
  OK column 6-12
Tof MS tab 16-5
TOF Resolution attribute 8-14
Tool Tray
  adding and removing tools 2-4
  description 2-3
  scroll buttons 2-4
toolbars
  introduction 2-2
  preferences button 2-2
  Query 14-2
  results browser 6-5
  Workflow Designer 7-4
tools
  adding and removing 2-4
  AutoMod Analysis 14-14–14-18
  BLAST Searching 13-6, 14-23–14-27
  Container Manager 5-2
  Databank Admin 13-2, 13-2–13-17
    description 13-2
  Databank Search 14-3–14-13
  De Novo Sequencing 14-19
  Digest Reagent tool 12-7–12-10
    description 12-7
  Expression Analysis 10-2
    description 10-2
  Gel Manager
    description 9-2
  Modifier tool 12-2–12-6
  Print tool 11-2–11-25
    description 11-2
  Sample Manager 4-2
    description 4-2
Troubleshooting
  installation on UNIX 1-20
  Linux 1-13
  UNIX 1-19
  Windows 1-6
Type databank attribute 13-4

U

uninstalling PLGS
  Linux 1-8
UNIX installation troubleshooting 1-20
Update Compression Type databank attribute 13-9
update current project 5-32, A-12, A-24
Update Renew Period databank attribute 13-10
Update URL Address databank attribute 13-9
updating projects 3-5
upregulation 10-11
upregulation filter 10-15
URL Address 13-8, 13-9
URL addresses E-2
URL Chooser dialog box 7-10
use replicate filter settings 10-14
user interface 2-2, 3-2

V

validate results 14-12, 14-16
variable modifications 14-11
view column
  sample lists 5-7
View Results parameter 5-23
viewing 6-34
  associated masses 6-34
  digest reagents 12-8
  exclude masses 6-34
  gel image 9-9
  gel results 9-9
  modifier reagents 12-3
  processed mass spectra 5-19

replicates for a cluster/protein 10-12
results 6-2
Expression experiment 10-10
sample annotation 9-10
sample information 5-23
sample lists 5-5
spectrum 5-19
workflows for clusters 10-12

W

Windows
executable file 1-4
installation 1-3
wizard
print 6-16
workflow
creating A-7, A-17
filters 7-11
for a cluster 10-12
icons 5-18
results 6-10, 6-12
templates 5-2, 5-6
adding 5-7, 5-20
attaching 5-20
printing 11-2
specifying 5-15
Workflow Designer 7-1–7-12
toolbar 7-4
workflow results 6-13
workpad
exclude masses 6-31
protein 6-27

X

x-axis
changing the view 6-20
range 6-24
scrolling 6-25
XML 2-20, 5-22, 14-5
importing project from 3-3
XSL style sheet 7-11, 7-12

Z

ZIP file
importing from 3-3
zoom view 6-25
zooming
gel image 9-9